

# **KDD con R.TeMiS**

## **su aplicación en el**

# **Proyecto Exhibitium**

José Pino Díaz

Universidad de Málaga, Andalucía Tech  
Escuela Técnica Superior de Ingeniería Industrial  
Campus de Teatinos, 29071  
Málaga, España

# **INTRODUCCIÓN**

# Proyecto Exhibitium

[Inicio](#) [Proyecto](#) [Objetivos](#) [Documentación](#) [Equipo](#) [Agenda](#) [Blog](#) [Contacto](#) [🔍](#)

EXHIBITIUM



EXHIBITIUM ES UN PROYECTO DE INVESTIGACIÓN  
FINANCIADO POR LA FUNDACIÓN BBVA

**Proyecto:** Generación de conocimiento sobre exposiciones artísticas temporales para su reutilización y aprovechamiento multivalente.

**Investigadora principal:** Nuria Rodríguez Ortega (Universidad de Málaga)

**Grupo de investigación:** iArHis\_Lab

**Categoría:** Humanidades Digitales

**Ayuda Fundación BBVA:** 60.000 Euros

**Web:** <http://exhibitium.com/>

# Expofinder

Sistema de información para el almacenamiento de referencias sobre exposiciones temporales de contenido artístico con sistema automático de captura de información y control de calidad.

**Expofinder (2015)** ha sido desarrollado por Antonio Cruces-Rodríguez, Profesor de la Universidad de Málaga.

Algunas cifras:

- 5.868 exposiciones
- 17.657 personas (autores y curadores)
- 11.758 entidades.

<http://www.expofinder.es/>

Sistema de gestión de datos ExpoFinder  
Plan de desarrollo de software y documentación técnica

V1.1 2016-01-12  
PDSDT-0101\_DEF.docx REV: 16/1/16

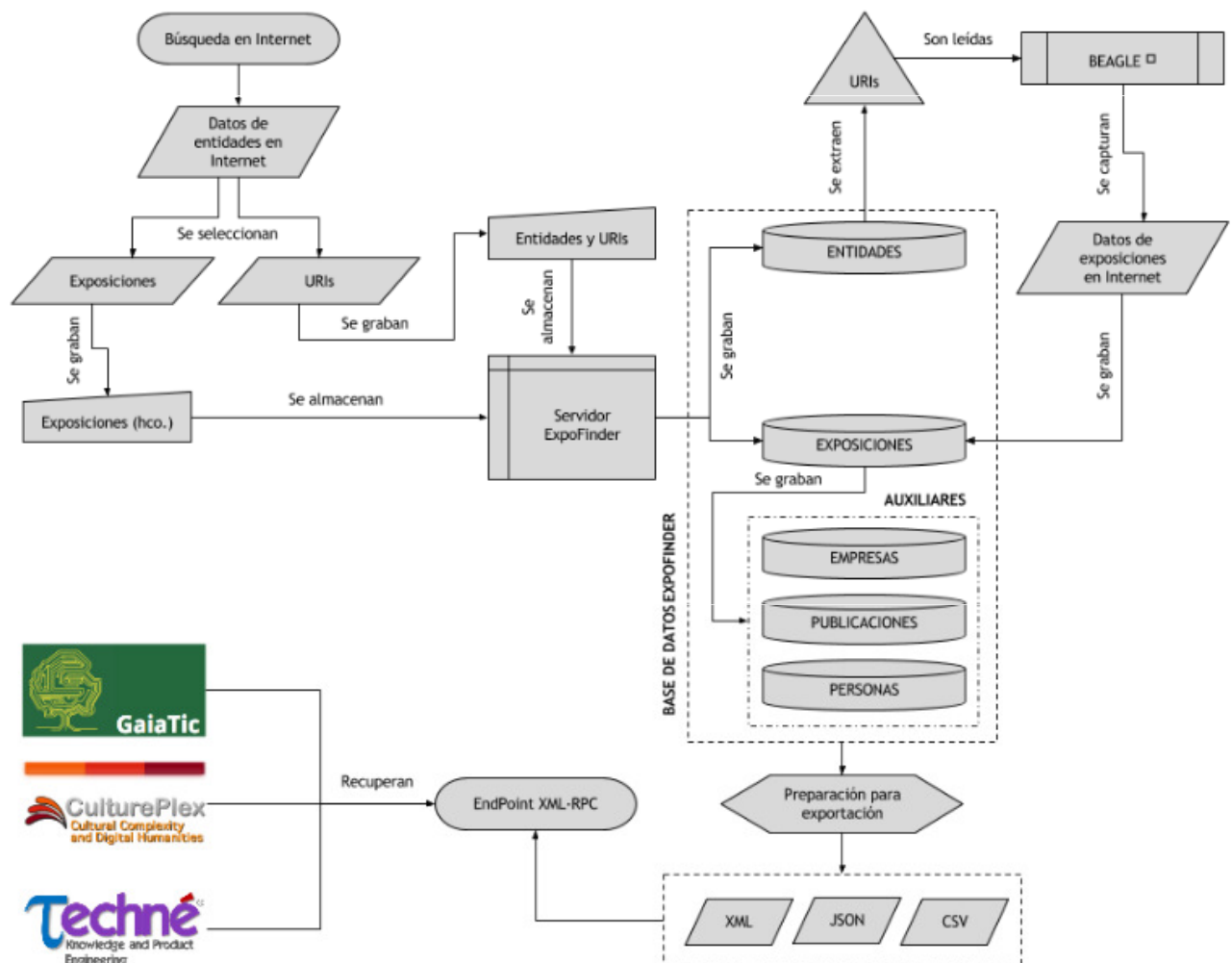
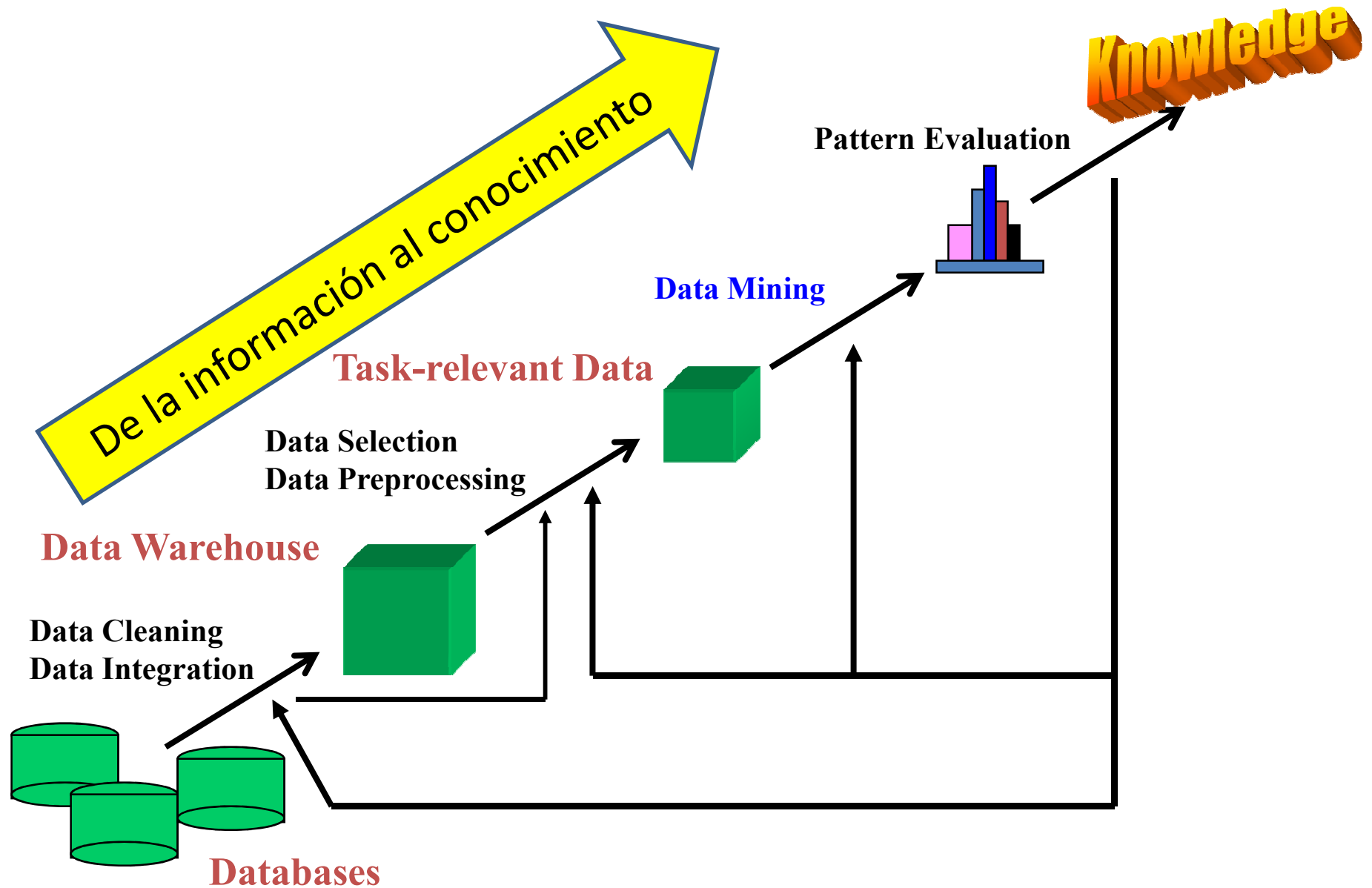


Ilustración 14. Relaciones entre los 3 sistemas principales de EF

# Knowledge Discovery Databases



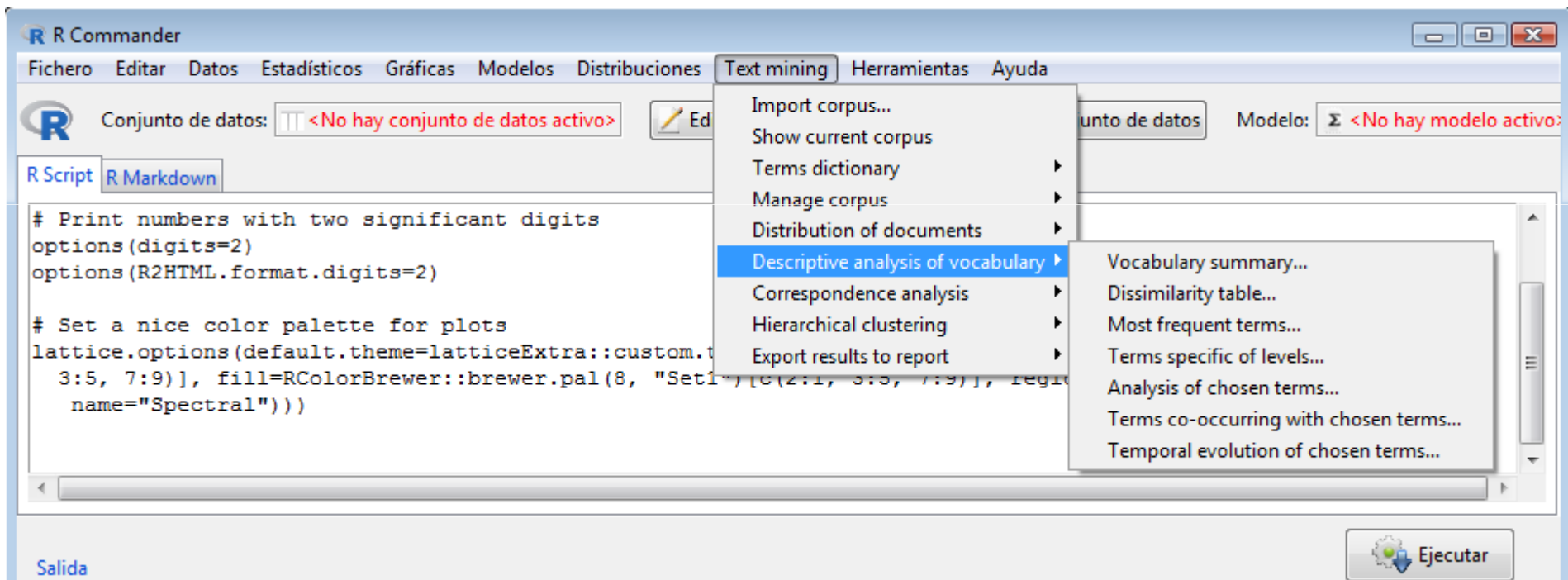
# **ANÁLISIS DEL CORPUS**

# R.Text Mining Solution

Bouchet-Valat, M., & Bastin, G. (2013). RcmdrPlugin.temis, a Graphical Integrated Text Mining Solution in R. *The R Journal* , 188-196.

Bouchet-Valat, M. (2016). *Package RcmdrPlugin.temis*.  
<https://cran.r-project.org/web/packages/RcmdrPlugin.temis/>

Garnier, B. (2014). *R.TeMiS. Une approche intégrée et libre de l'analyse de données textuelles*. <http://rtemis.hypotheses.org/>



- 1. Visualización del corpus activo y de los diccionarios de términos.**
- 2. Gestión y distribución del corpus.**
  1. Creación y análisis de subcorpus por términos o por variables
  2. Creación de tablas y gráficos de distribución.
- 3. Análisis descriptivo del léxico.**
  1. Resumen cuantitativo del vocabulario de términos.
  2. Tabla de disimilaridad.
  3. Términos más frecuentes.
  4. Términos específicos por modalidades de la variable.
  5. Análisis de términos concretos.
  6. Términos que coocurren con otros concretos.
  7. Evolución temporal de términos concretos.
- 4. Análisis factorial de correspondencias.**
  1. AFC de la matriz documentos-términos sin agregar ninguna variable.
  2. AFC de la tabla lexical completa agregando variables.
- 5. Clasificación ascendente jerárquica.**



# Corpus de estudio: 3391 exposiciones

	A	B	C	D	E	F	G	H	I
1	ENT_ID	ENT_TITL	EXH_ID	EXH_TITL	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh
2	1040	Museo de	68757	Bartolozzi	Calle de Santo Domín	67322: Ra	68748: Ba	42.819	
3	1075	Museo Ev	25042	Vagabundo	Camino Cabueñes, 2	25404: Evaristo Valle	43.533		
4	1075	Museo Ev	58543	Forma y O	Camino Cabueñes, 2	23241: Consuelo Valli	43.533		
5	1075	Museo Ev	65636	LIQUID, F	Camino Cabueñes, 3	67624: Michael Duney	43.534		
6	1075	Museo Ev	65638	Faena de	Camino C	1209: Mus	25404: Evaristo Valle	43.533	
7	1077	Museo Ba	47777	El oficio d	Calle Trinidad, 17, 33	25459: César Lucas	43.544		
8	1077	Museo Ba	49504	Silencios	Calle Trinidad, 17, 33	49505: Mingo Venero	43.544		
9	1085	Museo Etr	32887	Exposició	LLN-8, 132, 33509 Lla	33074: Cordelia Pickfo	43.412		
10	1156	Museo Mu	21523	27 negros	Rúa de Ramón del V	24337: Antonio Gonzá	42.883		
11	1156	Museo Mu	22068	Transform	Mallou Abaixo, 2, 157	24566: Vio	24592: Vio	42.892	
12	1156	Museo Mu	22934	De las tier	Praza Maior, 2, 32005	25731: Natasha Leler	42.335		
13	1156	Museo Mu	23062	Exposició	Rúa do Vilar, 49-51, 1	22113: Cristina Iglesia	42.878		
14	1156	Museo Mu	24071	Antonio M	Rúa Montero Ríos, 10	27022: Antonio Massó	42.240		
15	1156	Museo Mu	24112	La conser	LU-P-1102, 6, 27259	0	23812: Colectivo sin d	43.161	
16	1156	Museo Mu	24113	Il Cammin	Plaza Puerta del Sol, 1	27028: Steve McCurry			
17	1156	Museo Mu	24983	James W	Rúa de Ramón del V	26772: James Wellino	42.883		
18	1156	Museo Mu	24984	Curculio B	Rúa de Ramón del V	26393: Mark Manders	42.883		
19	1156	Museo Mu	25123	A orde da	Rúa de Ramón del V	27185: Carlos León	42.883		
20	1156	Museo Mu	25551	Océano	Av. da Atlántida, 160,	25690: Antón Patiño	42.224		
21	1156	Museo Mu	33188	La Ría de	Praza de	28487: Ge	28487: Germán Cruce	42.738	
22	1156	Museo Mu	43294	La Reboti	Rúa Padre Amoedo	0	22108: Eduardo Chillie	42.433	
23	116	Museu Aro	50171	Cantigas	Carrer Sant Bartolomé	0	51168: Juan Fernánde	38.484	
24	116	Museu Aro	61486	Made in A	Av. del Almte, Julio G		55293: [Sin determina	38.340	
25	1178	Museo Art	21695	Coco Cha	Calle Patio Chico, 2,	25653: Alberto Giaco	40.959		
26	1197	Museo Aro	7622	Cerámica	Calle Dr. Cervia Cabr	23812: Colectivo sin d	28.463		
27	1198	Museo de	7626	Figuras p	Calle Dr. Cervia Cabr	7629: Christine Mathie	28.463		
28	1199	Centro de	25102	Collages	Avenida de San Seb	25120: Jorge Rodrigu	28.463		
29	1201	TEA Tene	21252	Warags	Av. de San Sebastián	81994: Beatriz Rodrig	28.463		
30	1209	Museo de	48036	Sotileza /	Calle Rampa Sotileza	29276: Concha García	43.459		
31	1210	Museo de	56733	Una arque	Calle Bailén, 4, 39004	55293: [Sin determina	43.462		
32	1212	Museo Na	56209	De Bison	CA-134, 39360 Santill	23812: Colectivo sin d	43.377		
33	1212	Museo Na	62965	Elle n'étai	CA-134, 39360 Santill	63056: Sara Morante	43.377		
34	1223	Museo de	33197	La Real F	Paseo Poçillo, 1, 40100	La Granja de S. Ildet	40.902		
35	1240	Museo de	64312	El compá	Calle Manuel Rojas	64910: Ángel Idígoras	37.388		
36	1240	Museo de	64313	Camino e	Calle Manuel Rojas	64384: Mauro Di Gior	37.388		
37	136	Museo Ca	45185	Selección	Calle Navarro Rodrig	81256: Rafael Gadea	36.839		
38	136	Museo Ca	58562	Roscos	A-334, 04860 Olula de	66618: Pablo García	37.362		
39	1371	Museo Aro	48138	Exposició	Carrer Major, 11, 46320	63838: Lluís Bertrán	38.899		

Import Corpus

Load corpus from:

☐ Directory containing plain text files

☒ Spreadsheet file (CSV, XLS, ODS...)

☐ Factiva XML or HTML file(s)

☐ LexisNexis HTML file(s)

☐ Europresse HTML file(s)

☐ Alceste file(s)

☐ Twitter search

Language of texts in the corpus:

Español (es)

File encoding:

detect automatically

Text splitting:

☐ Split texts into smaller documents

Size of new documents:

1

paragraphs

Text processing:

☒ Ignore case

☐ Remove punctuation

☐ Remove digits

☐ Remove stopwords

☐ Apply stemming

☐ Edit stemming manually

Ayuda

Aceptar

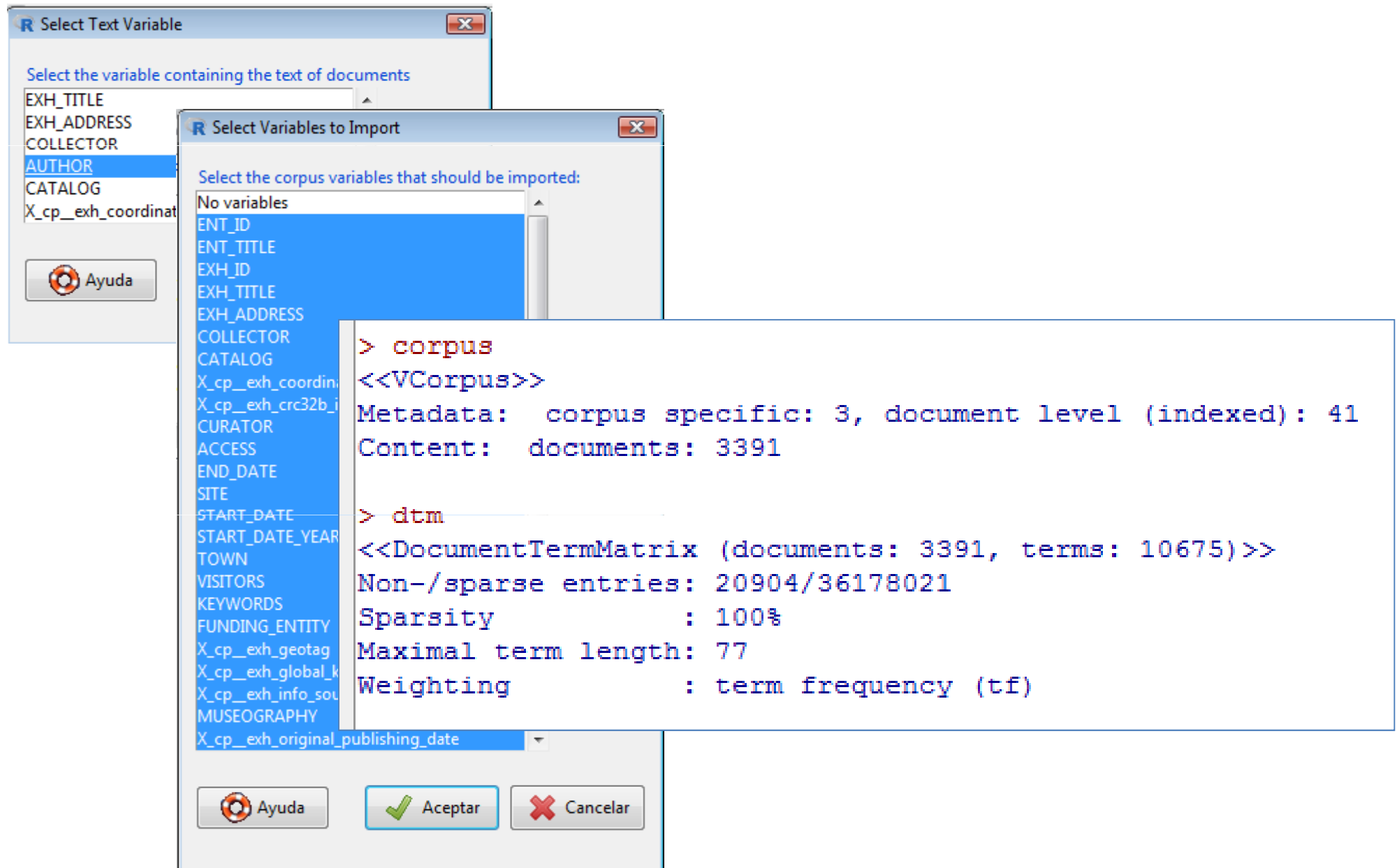
Cancelar

Q		
cp_exh	cp	
< 100		
< 100	arte	
< 100	mus	
< 100	mus	
< 100	mus	
< 100		
< 100		
< 100	arte	
< 100	arte	
< 100	arte	
< 100	arte	
< 100	arte	
< 100	arte	
< 100	exp	
< 100	arte	
< 100	arte	
< 100	arte	
< 100	artis	
< 100	artis	
< 100	artis	
< 100	pint	
< 100	artis	
< 100	mus	
100 a 499	pint	
< 100		
< 100		
< 100	arte	
< 100	arte	
< 100	arte	
< 100	mus	
< 100	arte	
< 100	arte	
< 100		
< 100	artis	
< 100	mus	
< 100		

# Variables de texto analizadas

- Autores de las obras
- Lugares de exposición: museos, galerías, salas de arte, etc.
- Comisarios de las exposiciones o curadores
- Coleccionistas
- Entidades organizadoras
- Entidades Financiadoras

# Selección de la variable de texto y var. suplementarias



The image shows two overlapping RStudio dialog boxes. The background dialog, 'Select Text Variable', prompts the user to 'Select the variable containing the text of documents' and lists variables including EXH\_TITLE, EXH\_ADDRESS, COLLECTOR, AUTHOR (highlighted), CATALOG, and X\_cp\_exh\_coordinat. The foreground dialog, 'Select Variables to Import', prompts the user to 'Select the corpus variables that should be imported:' and lists a long list of variables. The 'No variables' option is selected in this dialog. To the right of the foreground dialog, a terminal window displays the following R output:

```
> corpus
<<VCorpus>>
Metadata: corpus specific: 3, document level (indexed): 41
Content: documents: 3391

> dtm
<<DocumentTermMatrix (documents: 3391, terms: 10675)>>
Non-/sparse entries: 20904/36178021
Sparsity           : 100%
Maximal term length: 77
Weighting           : term frequency (tf)
```

At the bottom of the foreground dialog are three buttons: 'Ayuda' (with a lifebuoy icon), 'Aceptar' (with a green checkmark icon), and 'Cancelar' (with a red X icon).

# **EJEMPLOS DE RESULTADOS DEL ANÁLISIS DEL CORPUS**

# Ranking de autores:

## Listados alfabético y por ocurrencias

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Text Mining Analysis Results x +

file:///C:/Users/user/Desktop Buscar

### Text Mining Analysis Results

Corpus imported on NULL. Language: es.

Source: spreadsheet file C:/Users/user/Desktop/20160829\_EXPORTACIONES/20160829\_EXPORT\_PEPE\_VARIOS\_CRITERIOS\_SIN\_LIMITAR\_255.csv.

3391 documents and 10679 terms.

Processing options:

- Ignore case: enabled.
- Remove punctuation: disabled.
- Remove digits: disabled.
- Remove stopwords: disabled.
- Apply stemming: disabled.

**Terms dictionary sorted alphabetically**

	Occurrences
16189_peter_paul_rubens	42
16195_claudio_de_lorena	18
16196_pieter_brueghel	4
16200_josé_garcía_ramos	1
19461_gonzalo_bilbao	3
21592_luis_de_morales	7
21598_francisco_de_zurbarán	35
21605_anton_van_dyck	14
21610_francisco_de_goya_y_lucientes	71
21614_julio_romero_de_torres	10
21618_bartolomé_esteban_murillo	36
21784_rafael_esteve_y_vilella	1
21785_luigi_pellegrino_scaramuccia	1
21786_paolo_veronese	11
21788_pascual_pedro_moles	1
21821_hermanos_penicaud	1
21823_juan_de_valdés_leal	10
21949_alberto_durero	17

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Text Mining Analysis Results x +

file:///C:/Users/user/Desktop Buscar

### Text Mining Analysis Results

**Terms dictionary sorted by number of occurrences**

	Occurrences
55293_[sin_determinar]	218
6664_pablo_ruiz_picasso	149
21610_francisco_de_goya_y_lucientes	71
23812_colectivo_sin_determinar	69
25924_antoni_tàpies_i_puig	60
7319_salvador_dalí	56
7677_joan_miró	55
16189_peter_paul_rubens	42
22131_antonio_saura	40
7675_diego_velázquez	40
25538_joaquín_sorolla_y_bastida	37
21618_bartolomé_esteban_murillo	36
21598_francisco_de_zurbarán	35
25463_equipo_crónica_o_crónicas_de_la_realidad	35
24401_ignasi_aballí	33
34097_josé_de_ribera_y_cucó	32
6527_el_greco	32
22112_luis_gordillo	31
25536_ignacio_pinazo_camarlench	31
24525_pablo_palazuelo_de_la_peña	30
27348_max_ernst	29
22103_eduardo_arroyo_rodríguez	27
22106_carmen_calvo	27
24436_antoni_muntadas	27
25223_eusebio_sempere	27
26582_robert_rauschenberg	27
27345_julio_gonzález	27
7264_tiziano	27
22113_cristina_iglesias_fernández_berrido	26
25477_juan_uslé	26
25854_txomin_badiola	26

# Ranking de curadores:

## Listados alfabético y por ocurrencias

Text Mining Analysis Results

Corpus imported on NULL. Language: es.

Source: spreadsheet file C:/Users/user/Desktop/20160829\_EXPORTACIONES/20160829\_EXPORT\_ENT\_FUENTE/20160829\_EXPORT\_PEPÉ\_copia\_SIN\_AUT\_SIN\_CONFIRMAR.csv.

3391 documents and 991 terms.

Processing options:

- Ignore case: enabled.
- Remove punctuation: disabled.
- Remove digits: disabled.
- Remove stopwords: disabled.
- Apply stemming: disabled.

**Terms dictionary sorted alphabetically**

	Occurrences
16170_teresa_posada	6
17851_manuel_arias_martínez	1
17852_miguel_ángel_marcos_villán	1
17853_antônio_filipe_pimentel	1
21773_andrés_úbeda_de_los_cobos	4
21983_david_martínez	1
21987_antonio_malpica_cuello	1
22100_francisco_calvo_serraller	8
22149_arte_contemporáneo_gacma	1
22178_ana_arambarri	1
23081_maría_josé_garcía_borge	1
23088_juan_manuel_garcía_ruiz	1
23100_rafael_levenfeld_ortiz	4
23103_valentín_vallhonrat	4
23263_concha_mayordomo	1
23282_maría_jesús_aragoneses	1
23465_félix_fanés	2

Text Mining Analysis Results

**Terms dictionary sorted by number of occurrences**

	Occurrences
27093_fernando_francés	52
25349_miguel_von_hafe	15
43029_joão_fernandes	15
43687_guillermo_solana	15
49387_salvador_carretero	14
60645_blanca_de_la_torre	14
77401_lynne_cooke	13
49388_isabel_portilla	12
50849_manuel_borja-villel	12
40636_tomás_llorens_serra	10
48555_ferran_barenblit	10
67643_bartomeu_marí_i_ribas	10
7378_josé_manuel_matilla_rodríguez	10
7867_javier_barón_thaidigsmann	10
42514_fernando_pérez_suescun	9
6874_manuela_b_mena_marqués	9
7361_alejandro_vergara_sharp	9
22100_francisco_calvo_serraller	8
26812_chus_martínez	8
55128_álvaro_rodríguez_fominaya	8
63830_fernando_castro_flórez	8
64652_juan_antonio_álvarez_reyes	8
25501_manuel_oliveira	7
41138_yolanda_romero	7
65038_manuel_segade	7
66035_josé_miguel_garcía_cortés	7
6854_miguel_falomir_faus	7
7474_montse_aguer	7
81019_belén_poole	7
16170_teresa_posada	6
23653_carmen_giménez_martín	6

# Ranking de entidades:

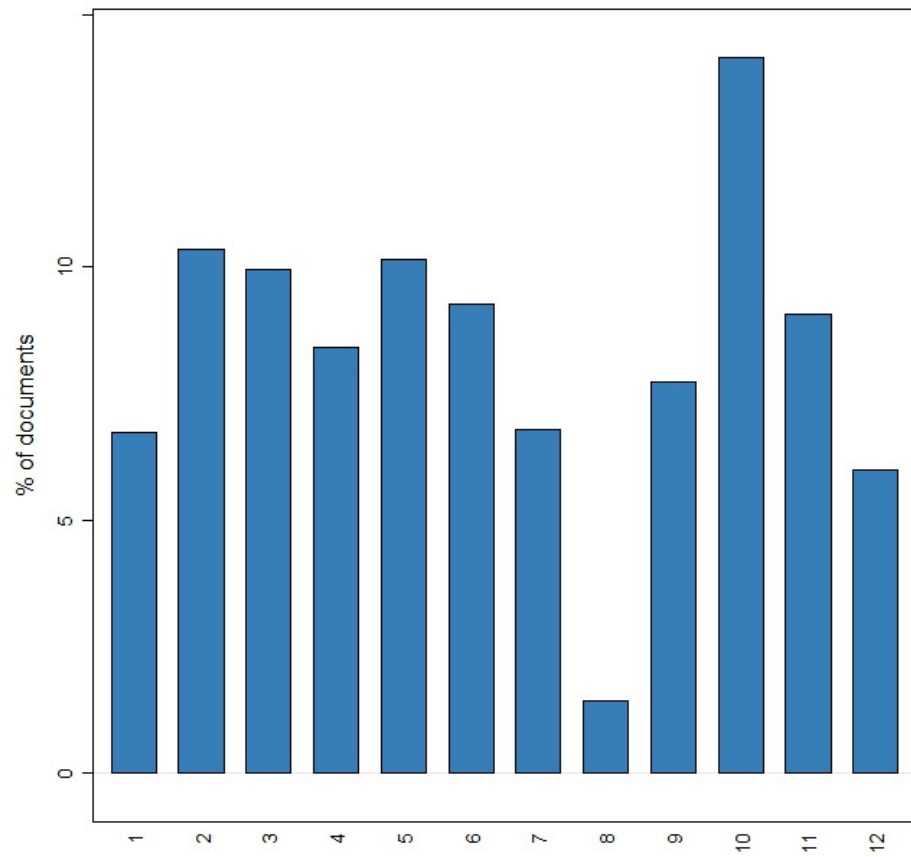
## Entidades que financian

Terms dictionary sorted by number of occurrences	
	Occurrences:
3530:_fundación_iberdrola	13
2188:_fundación_bbva	12
22099:_fundación_amigos_del_museo_del_prado	11
39926:_fundación_banco_sabadell	10
7376:_fundación_axa	10
58260:_el_periódico	9
64428:_corts_valencianes	7
64867:_heineken	7
2637:_ministerio_de_educación,_cultura_y_deporte	5
23505:_comunidad_de_madrid	4
73544:_fundación_abertis	4
2033:_fundacion_bancaja	3
63273:_terra_foundation_for_american_art_europe	3
1995:_centro_de_arte_contemporáneo_málaga	2
22176:_fundación_telefónica	2
264:_museu_d'art_contemporani_de_barcelona	2
26512:_acciona_s.a.	2
2668:_fundación_barrié_sede_a_coruña	2
27663:_generalitat_valenciana	2
29096:_lisson_gallery	2
34211:_baldwin_gallery	2
43919:_año_dual_españa_-_japón	2
45821:_iniciarte	2
77308:_kerabén	2
77614:_fomento_urbano_de_castellón_s.a	2
8176:_fundación_casa_ducal_de_medinaceli	2
839:_museo_nacional_de_artes_decorativas	2
1807:_consorci_de_museus_de_la_comunitat_valenciana._generalitat	1
1970:_fundación_cañada_blanck	1
2018:_la_nau,_centro_cultural_de_la_universitat_de_valència	1
2044:_caja_de_cultura_de_valencia	1

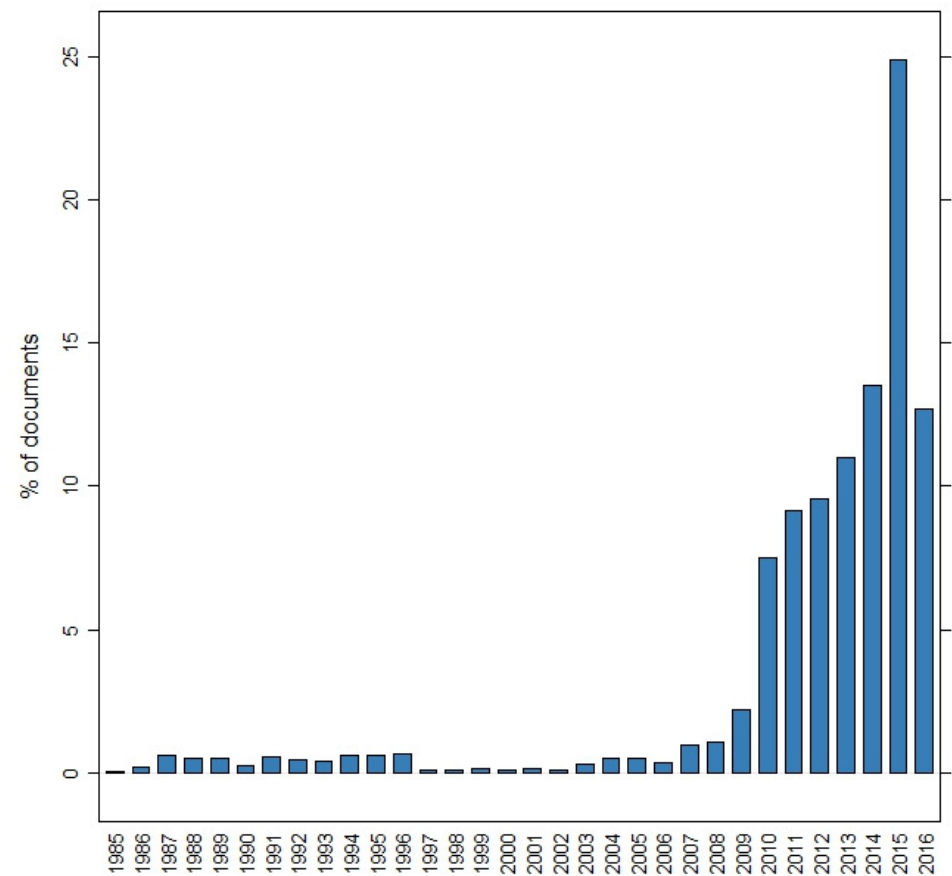


# Distribuciones de las exposiciones (Una variable)

Distribution of exhibitions by month in Spain



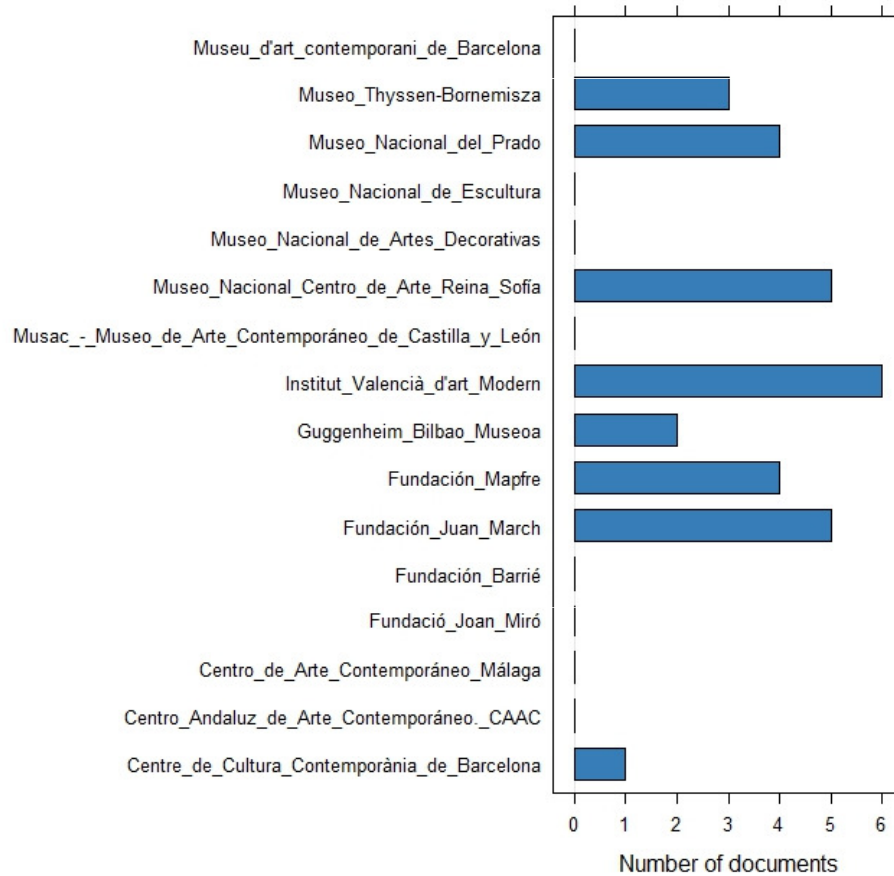
Distribution of exhibitions by year in Spain



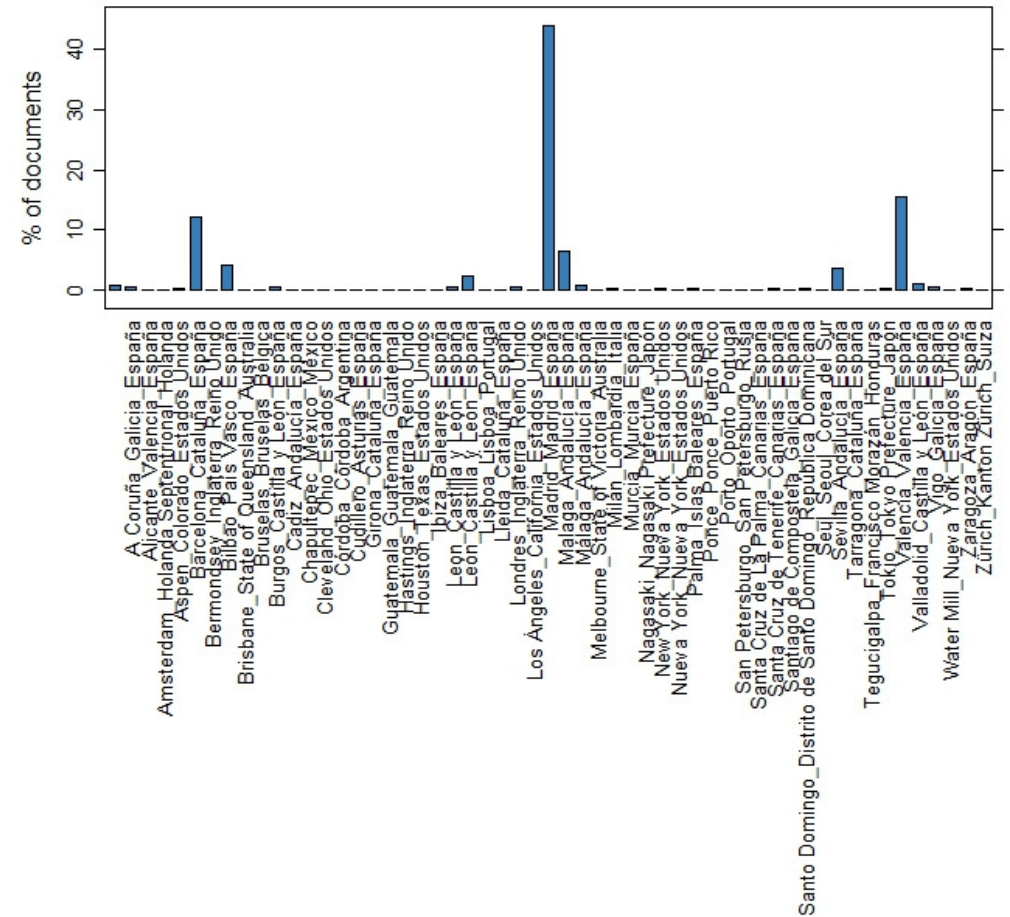


# Distribuciones de las exposiciones (Una variable)

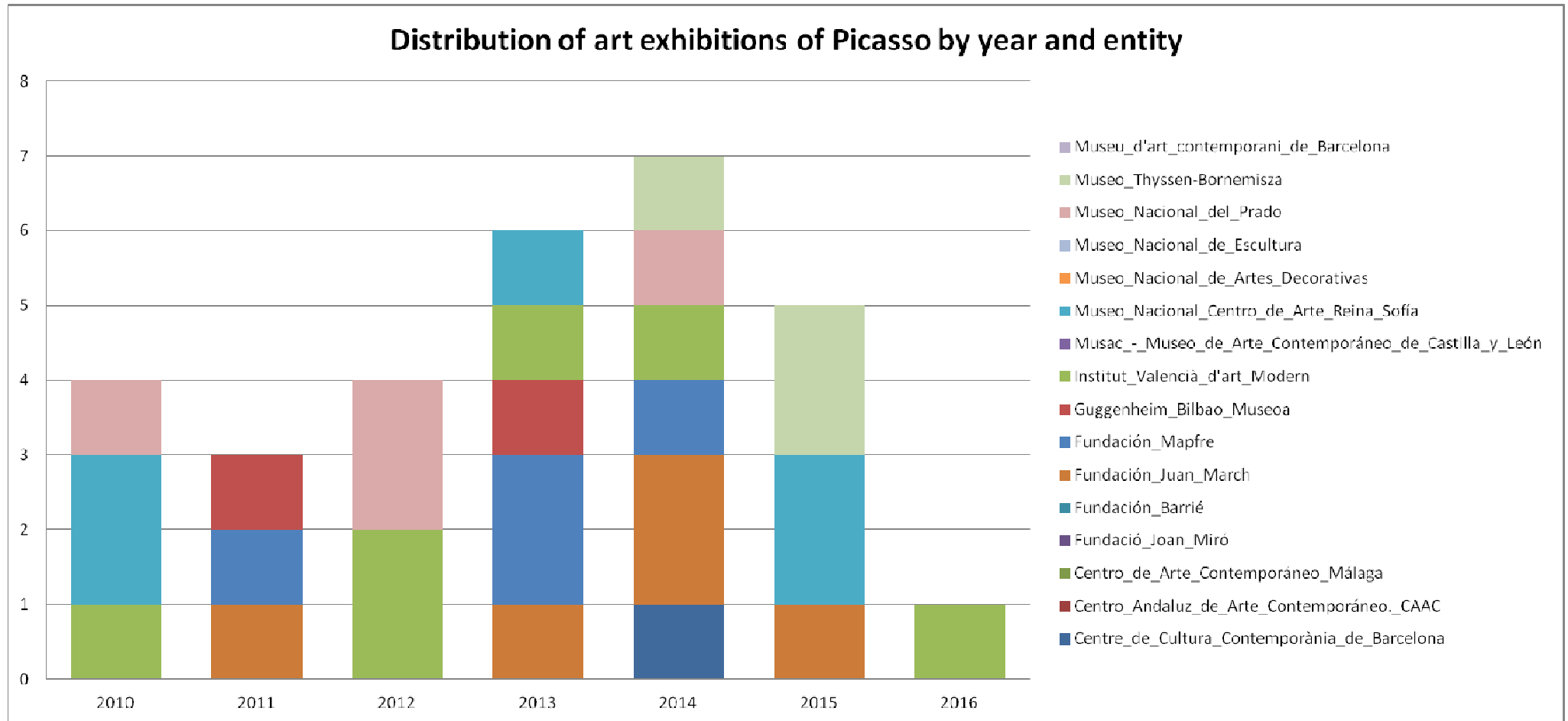
Distribution of documents by Entity



Distribution of documents by TOWN

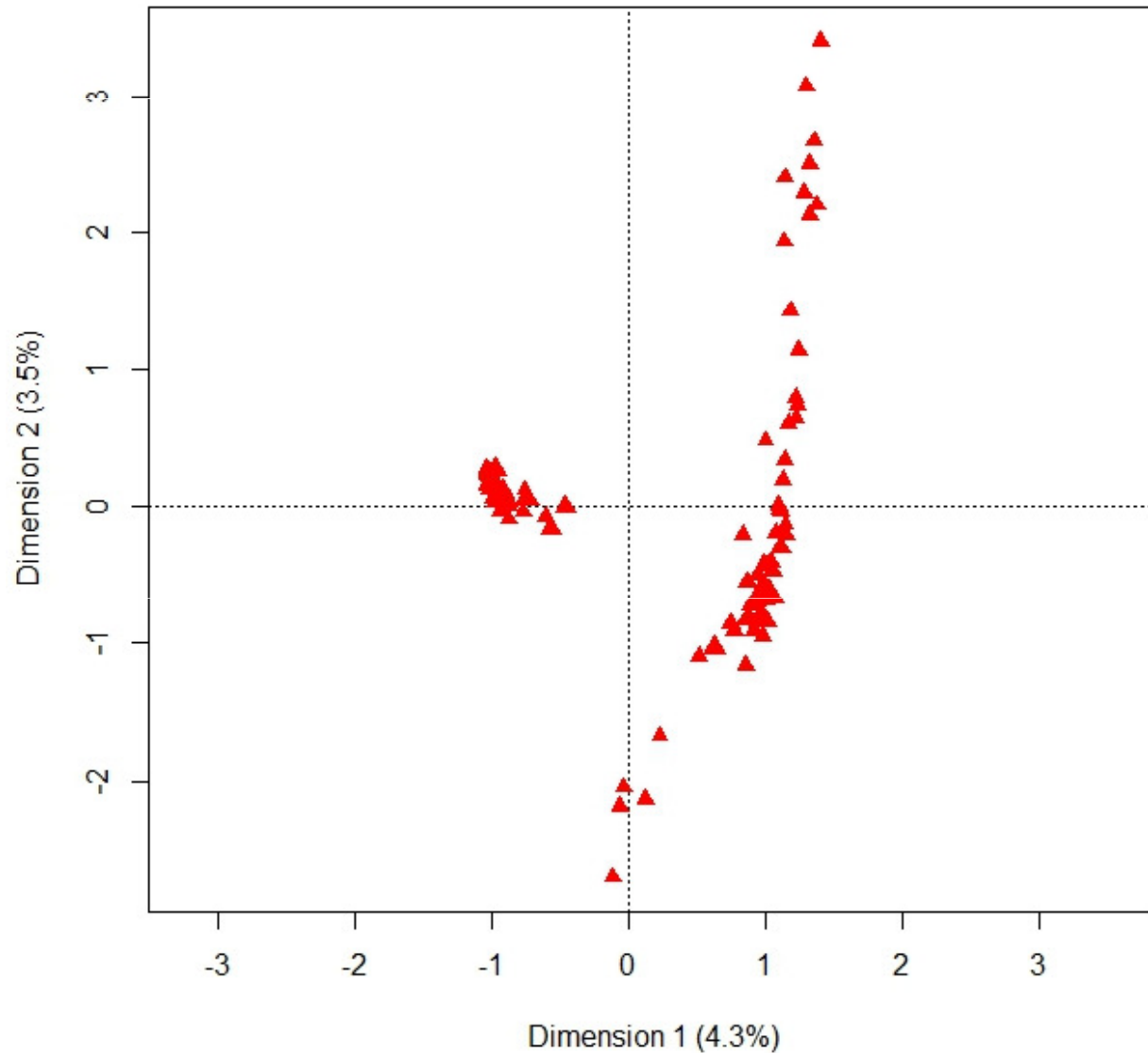


# Distribuciones de las exposiciones (Dos variables)

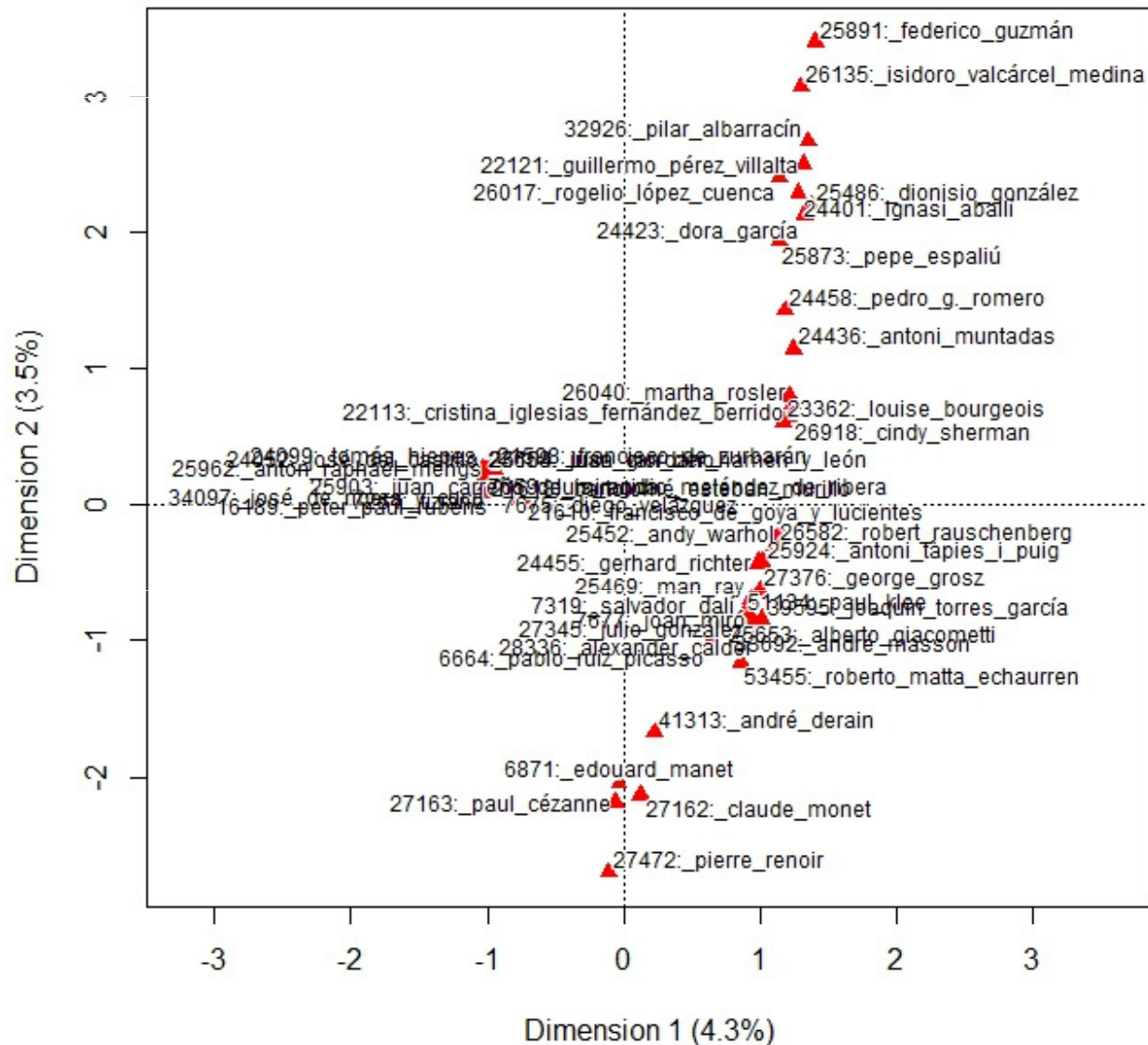


# Correspondence analysis:

## AC de la matriz exp/autores



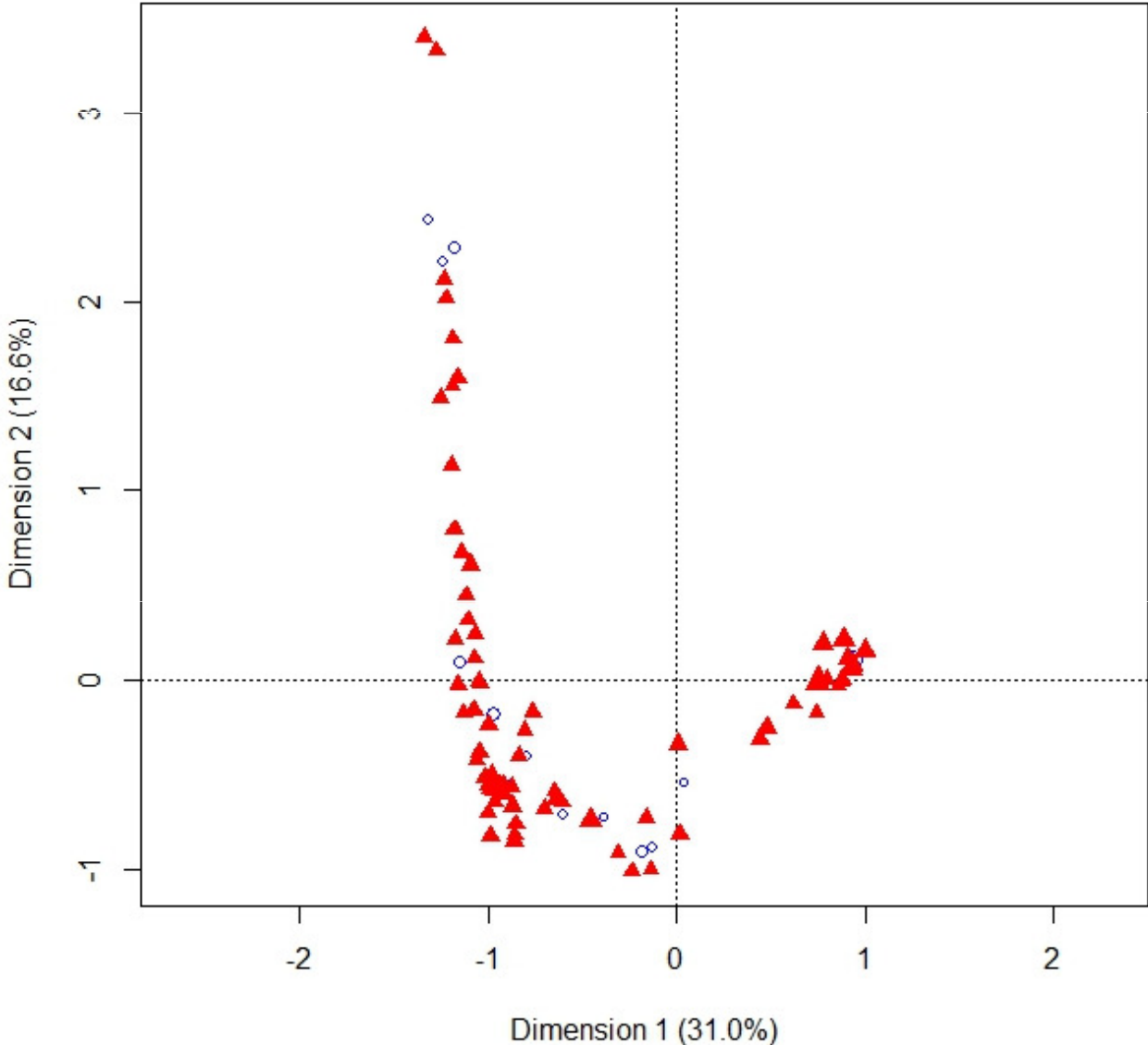
# Correspondence analysis: AC de la matriz exp/autores



# Correspondence analysis:

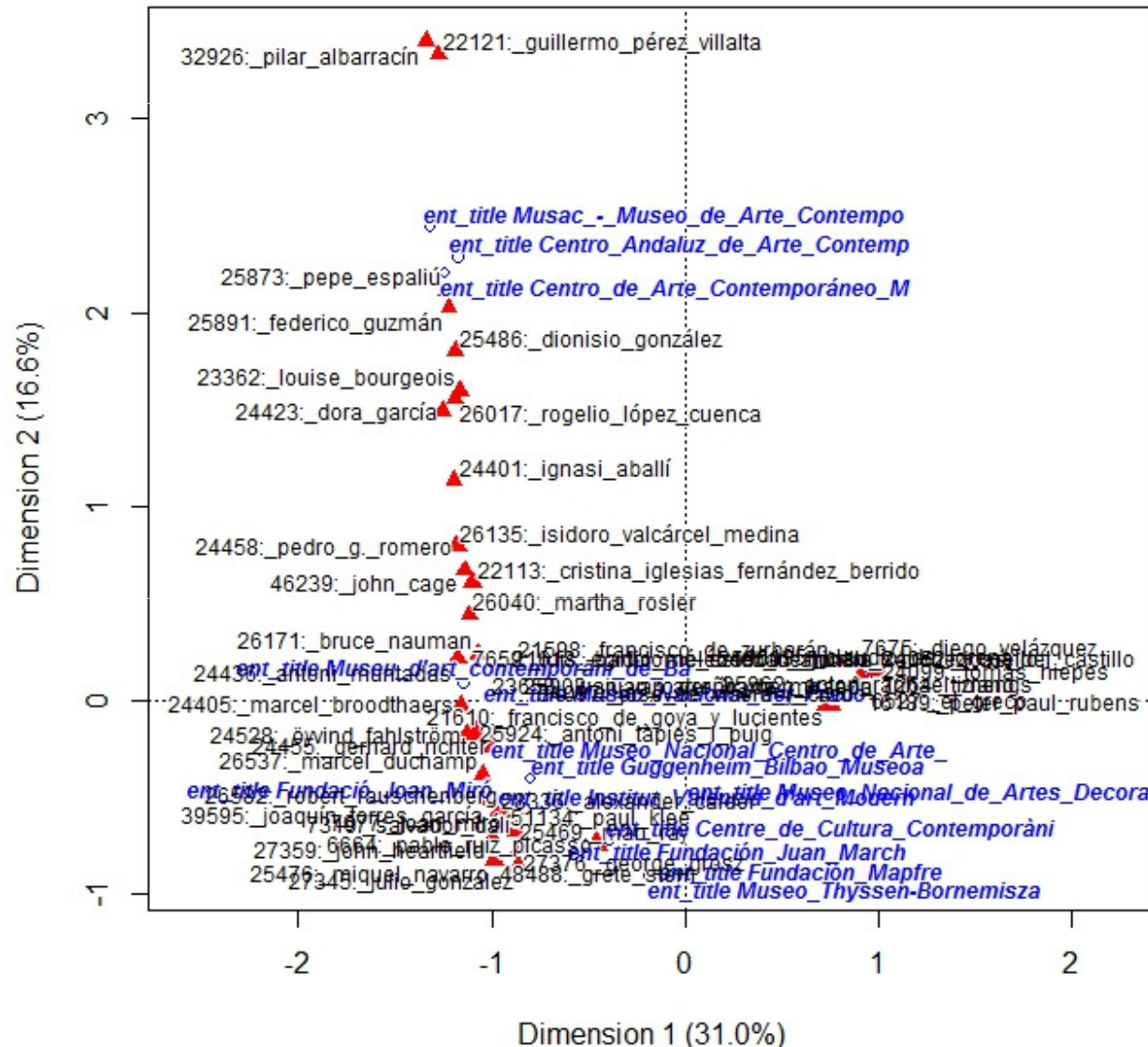
# AC de la matriz exp/aut

## agregando la variable museos



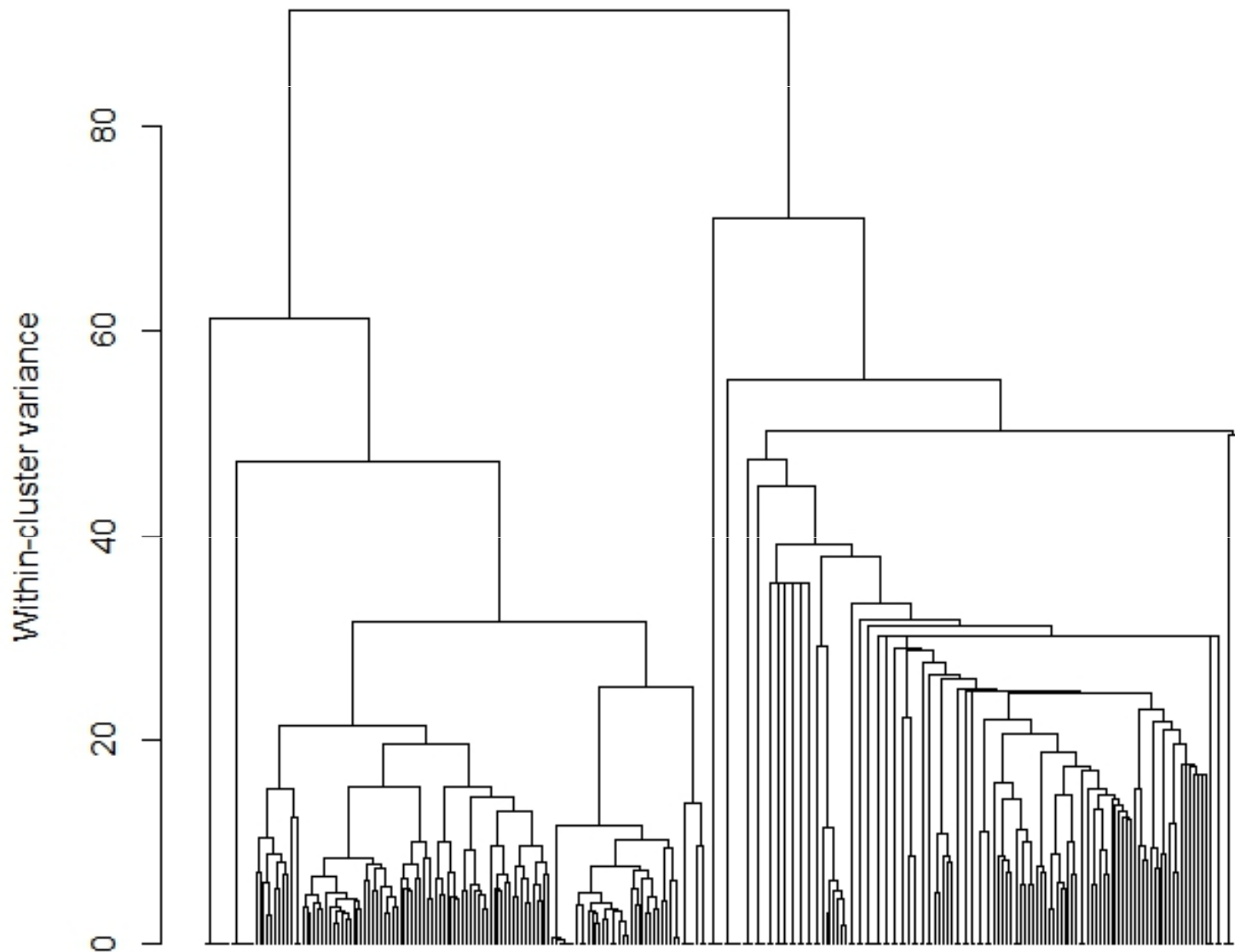
# Correspondence analysis:

## AC de la matriz exp/aut agregando la variable museos



# Hierarchical clustering:

## Clasificación de las exposiciones según su similaridad por autoría



# **EJEMPLOS DE RESULTADOS DEL ANÁLISIS DE DIFERENTES SUBCORPUS**

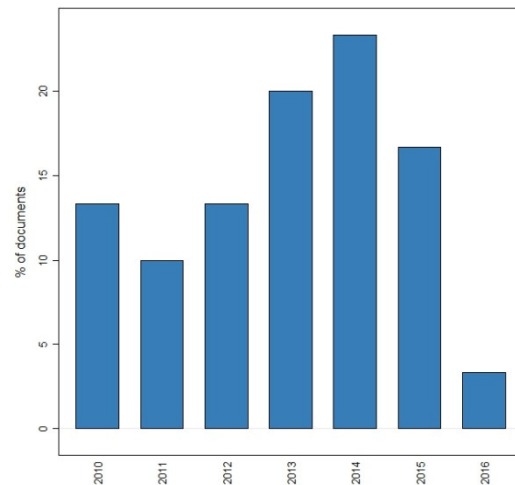


**SUBCORPUS PICASSO**

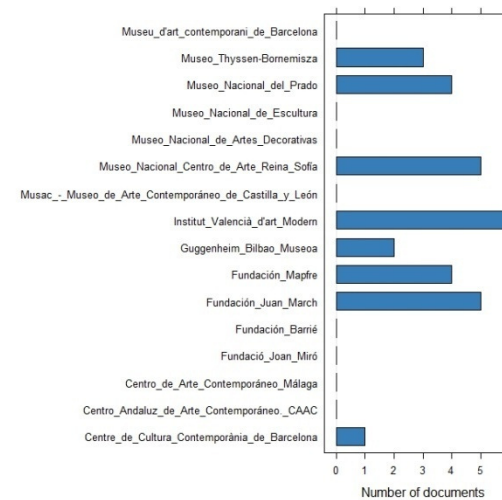
# Subcorpus Picasso

## Distribuciones, 1 y 2 variables

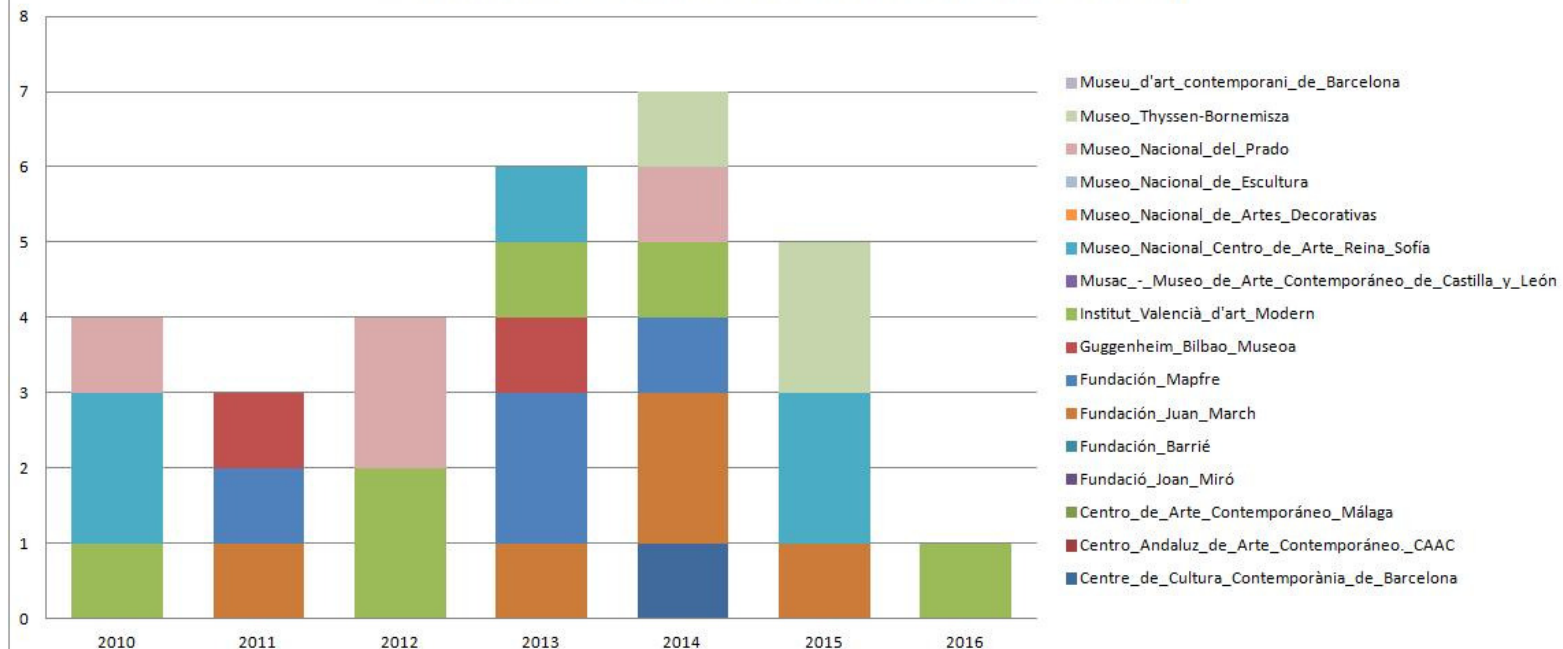
### Distribution of documents by year



### Distribution of documents by Entity

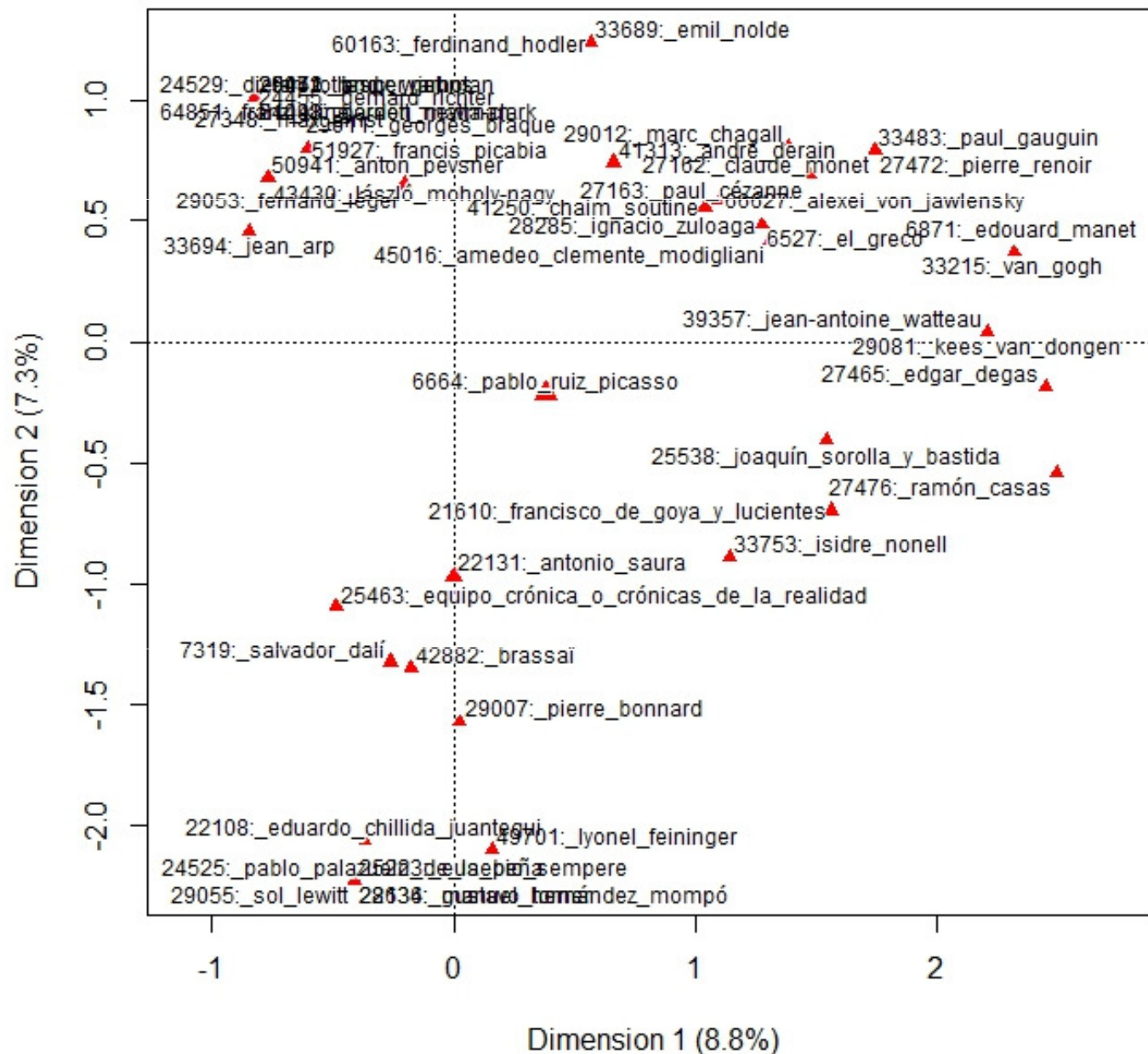


### Distribution of art exhibitions of Picasso by year and entity

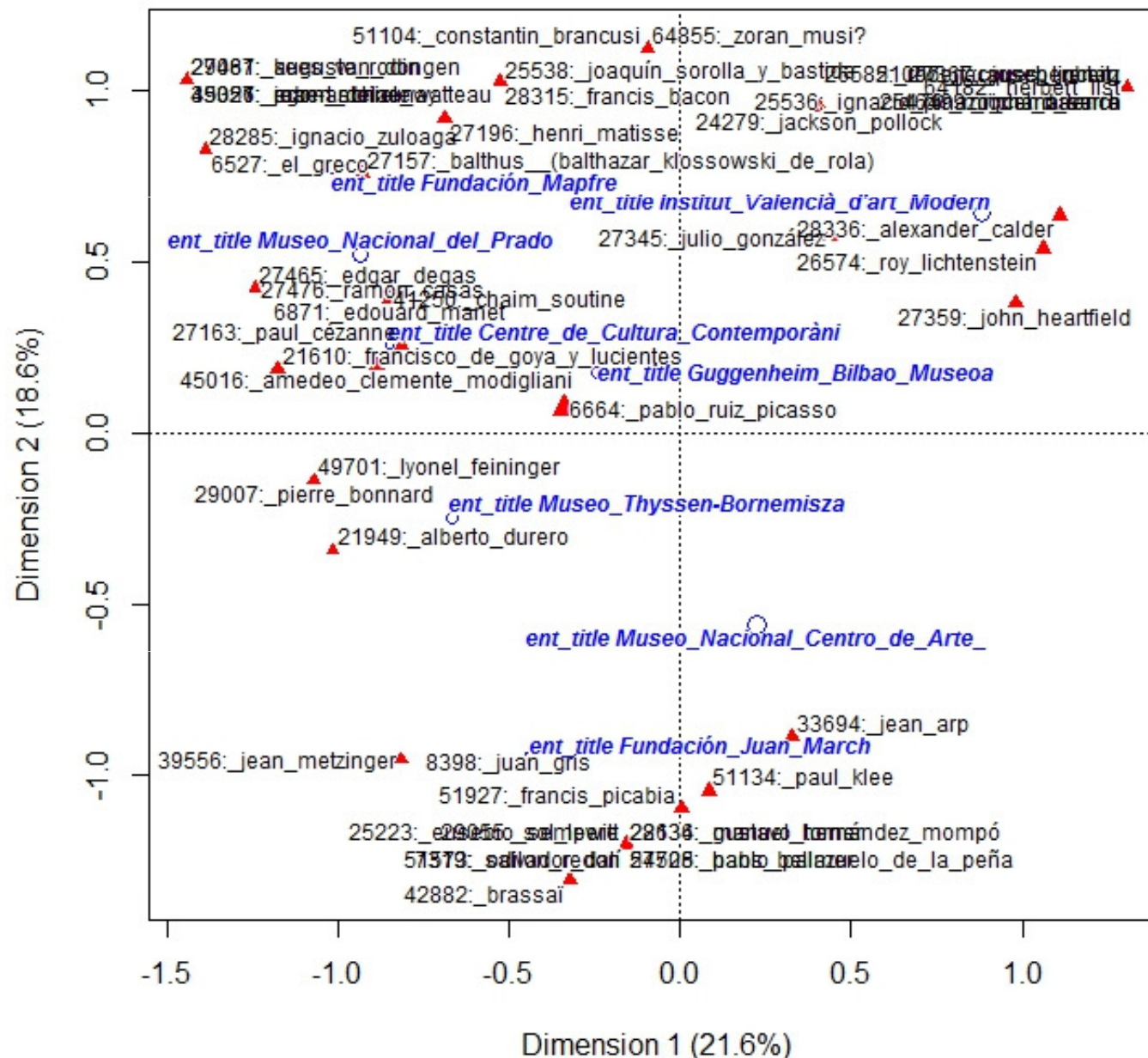


# Subcorpus Picasso

## AC matriz exp/autores



# AC de la matriz exp/aut con var. museos



# Subcorpus Picasso

## HC autores de las obras

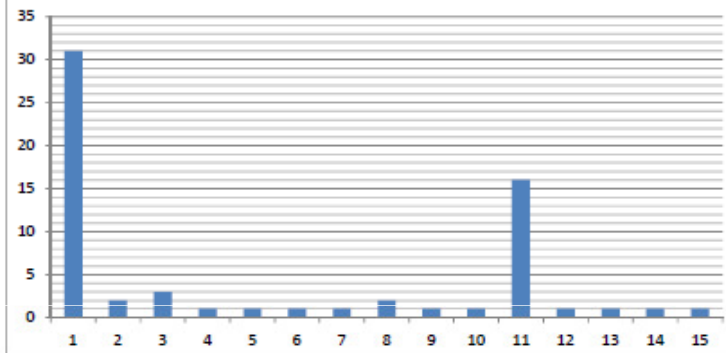
### « Picasso » subcorpus

Hierarchical clustering of 64 expositions and 130 authors

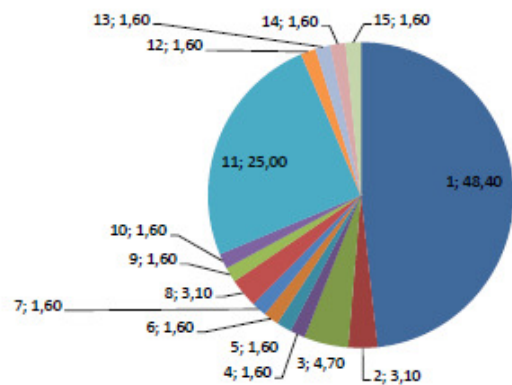
#### Clusters summary:

	1	2	3	4	5	6	7	8	9	10	11	12
Number of documents	31.0	2.0	3.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	16.0
% of documents	48.4	3.1	4.7	1.6	1.6	1.6	1.6	1.6	3.1	1.6	1.6	25.0
Within-cluster variance	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	5.4	0.0	0.0	6.8
	13	14	15									
Number of documents	1.0	1.0	1.0									
% of documents	1.6	1.6	1.6									
Within-cluster variance	0.0	0.0	0.0									

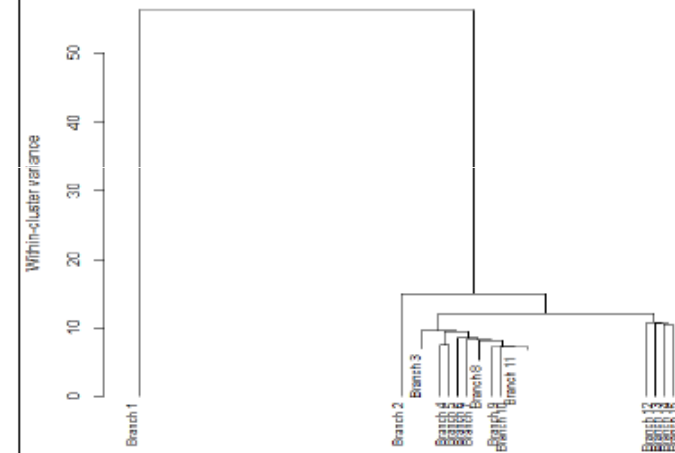
Number of expositions by cluster



% of expositions by cluster



Clusters dendrogram



Number of cluster	Number expositions by cluster	% of expositions	Terms specific of cluster
1	31	48,40	6664: pablo ruiz picasso
2	2	3,10	78435: honoré de balzac
3	3	4,70	47280: José ruiz y blasco
4	1	1,60	29074: théophile alexandre steinien 72774: francisco sancha lengo
5	1	1,60	27465: edgar degas 45016: amedeo clemente modigliani
6	1	1,60	19461: gonzalo bilbao 33713: José dominguez bécquer
7	1	1,60	24522: richard hamilton 7835: jean laurent y minier
8	2	3,10	27345: julio gonzález
9	1	1,60	25142: rafael canogar 33855: antonio lópez garcia 72004: francisco rojas 72005: fernando sordo
10	1	1,60	22106: camen calvo 22131: antonio saura 24447: wolf vostell 25453: manolo valdés blasco 25463: equipo crónica o crónicas de la realidad 39935: equipo realidad 64712: christopher makos
11	16	25,00	29011: georges braque 7319: salvador dali
12	1	1,60	33483: henry paul gauguin
13	1	1,60	25656: david douglas duncan
14	1	1,60	25473: xavier vilató 50563: j. fin
15	1	1,60	67657: ramon reventós



# Subcorpus Picasso

## HC autores de las obras

### Terms specific of cluster 8:

	% Term/Level	% Level/Term	Global %	Level	Global	t value	Prob.
27345:_julio_gonzález	29	100	0.87	2	2	3.2	0.0008
64794:_joan_gonzález	14	100	0.43	1	1	1.9	0.0304
40982:_antoni_gaudí_i_cornet	14	50	0.87	1	2	1.6	0.0601

Documents specific of cluster 8:

Chi2 dist. to centroid

94	5.3
2733	9.4

94

27345:\_Julio\_González 40982:\_Antoni\_Gaudí\_i\_Cornet 64794:\_Joan\_González  
6664:\_Pablo\_Ruiz\_Picasso

2733

27196:\_Henri\_Matisse 27345:\_Julio\_González 6664:\_Pablo\_Ruiz\_Picasso

### Terms specific of cluster 9:

	% Term/Level	% Level/Term	Global %	Level	Global	t value	Prob.
25142:_rafael_canogar	17	100	0.43	1	1	1.9	0.026
33855:_antonio_lópez_garcía	17	100	0.43	1	1	1.9	0.026
72004:_francisco_rojas	17	100	0.43	1	1	1.9	0.026
72005:_fernando_sordo	17	100	0.43	1	1	1.9	0.026

Documents specific of cluster 9:

Chi2 dist. to centroid

2215	0.0
------	-----

2215

25142:\_Rafael\_Canogar 33855:\_Antonio\_López\_García 6664:\_Pablo\_Ruiz\_Picasso  
72004:\_Francisco\_Rojas 72005:\_Fernando\_Sordo 8398:\_Juan\_Gris

### Terms specific of cluster 10:

	% Term/Level	% Level/Term	Global %	Level	Global	t value	Prob.
22106:_carmen_calvo	12	100	0.43	1	1	1.8	0.035
22131:_antonio_saura	12	100	0.43	1	1	1.8	0.035
24447:_wolf_vostell	12	100	0.43	1	1	1.8	0.035
25453:_manolo_valdés_blasco	12	100	0.43	1	1	1.8	0.035
25463:_equipo_crónica_o_crónicas_de_la_realidad	12	100	0.43	1	1	1.8	0.035
39935:_equipo_realidad	12	100	0.43	1	1	1.8	0.035
64712:_christopher_makos	12	100	0.43	1	1	1.8	0.035

Documents specific of cluster 10:

Chi2 dist. to centroid

### Terms specific of cluster 11:

	% Term/Level	% Level/Term	Global %	Level	Global	t value	Prob.
29011:_georges_braque	3.6	100	2.2	5	5	1.4	0.078
7319:_salvador_dali	3.6	100	2.2	5	5	1.4	0.078
6664:_pablo_ruiz_picasso	11.5	25	27.8	16	64	-6.7	0.000
42280:_josé_ruiz_y_blasco	0.0	0	1.3	0	3	-1.5	0.061

Documents specific of cluster 11:

Chi2 dist. to centroid

741	10.6
1419	11.5
2255	11.8
2749	12.1
24	14.5
441	15.0
1309	15.1
2085	15.2
2729	15.5
475	15.6
1305	15.9
1327	16.1
526	16.5
827	17.9
486	20.8
2727	21.4

741

26537:\_Marcel\_Duchamp 29011:\_Georges\_Braque 29053:\_Fernand\_Léger  
39554:\_Albert\_Gleizes 39556:\_Jean\_Metzinger 41313:\_André\_Derain 43507:\_Jacques\_Villon  
51927:\_Francis\_Picabia 6664:\_Pablo\_Ruiz\_Picasso 78043:\_Marie-Laurencin 8398:\_Juan\_Gris

1419

25617:\_Gino\_Severini 27049:\_María\_Blanchard 27196:\_Henri\_Matisse  
29011:\_Georges\_Braque 29053:\_Fernand\_Léger 29069:\_Diego\_Rivera  
51097:\_Jacques\_Lipchitz 51099:\_Henri-Laurens 6664:\_Pablo\_Ruiz\_Picasso 8398:\_Juan\_Gris

2255

29011:\_Georges\_Braque 6664:\_Pablo\_Ruiz\_Picasso

2749

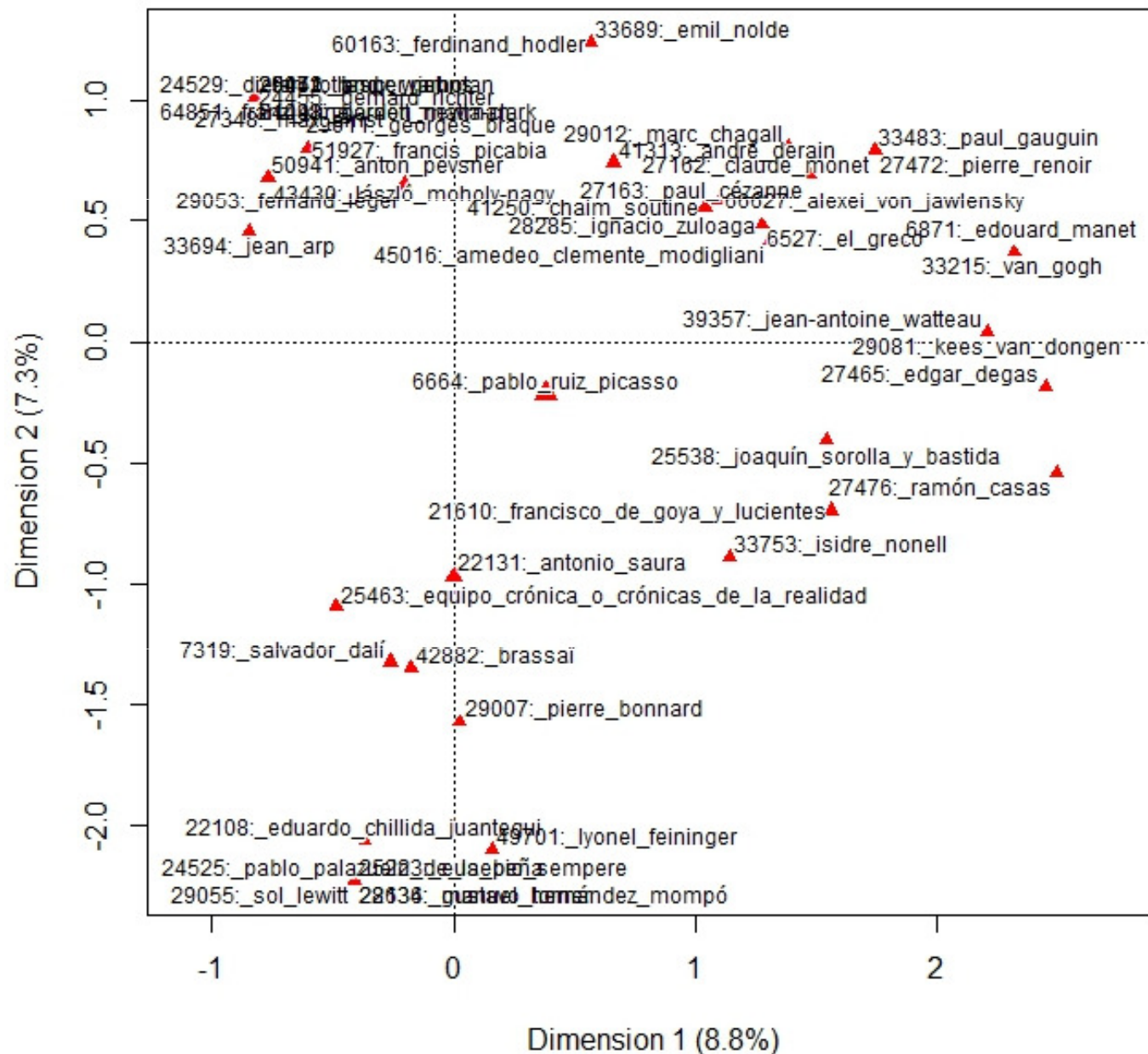
25449:\_Joaquín\_Peinado 27049:\_María\_Blanchard 29011:\_Georges\_Braque  
33375:\_Paul\_Strand 41313:\_André\_Derain 43505:\_Federico\_García\_Lorca  
43506:\_José\_Moreno\_Villa 43507:\_Jacques\_Villon 6664:\_Pablo\_Ruiz\_Picasso  
7677:\_Joan\_Miró

24

25653:\_Alberto\_Giacometti 28359:\_Óscar\_Domínguez 50352:\_Josep\_Maria\_Sert  
6664:\_Pablo\_Ruiz\_Picasso 69001:\_Apel·les\_Fenosa 7319:\_Salvador\_Dali  
81009:\_Kollar\_François 81012:\_Demetre\_Chiparus 81013:\_Coco\_Chanel 8398:\_Juan\_Gris

# Subcorpus Picasso

## AC matriz exp/autores

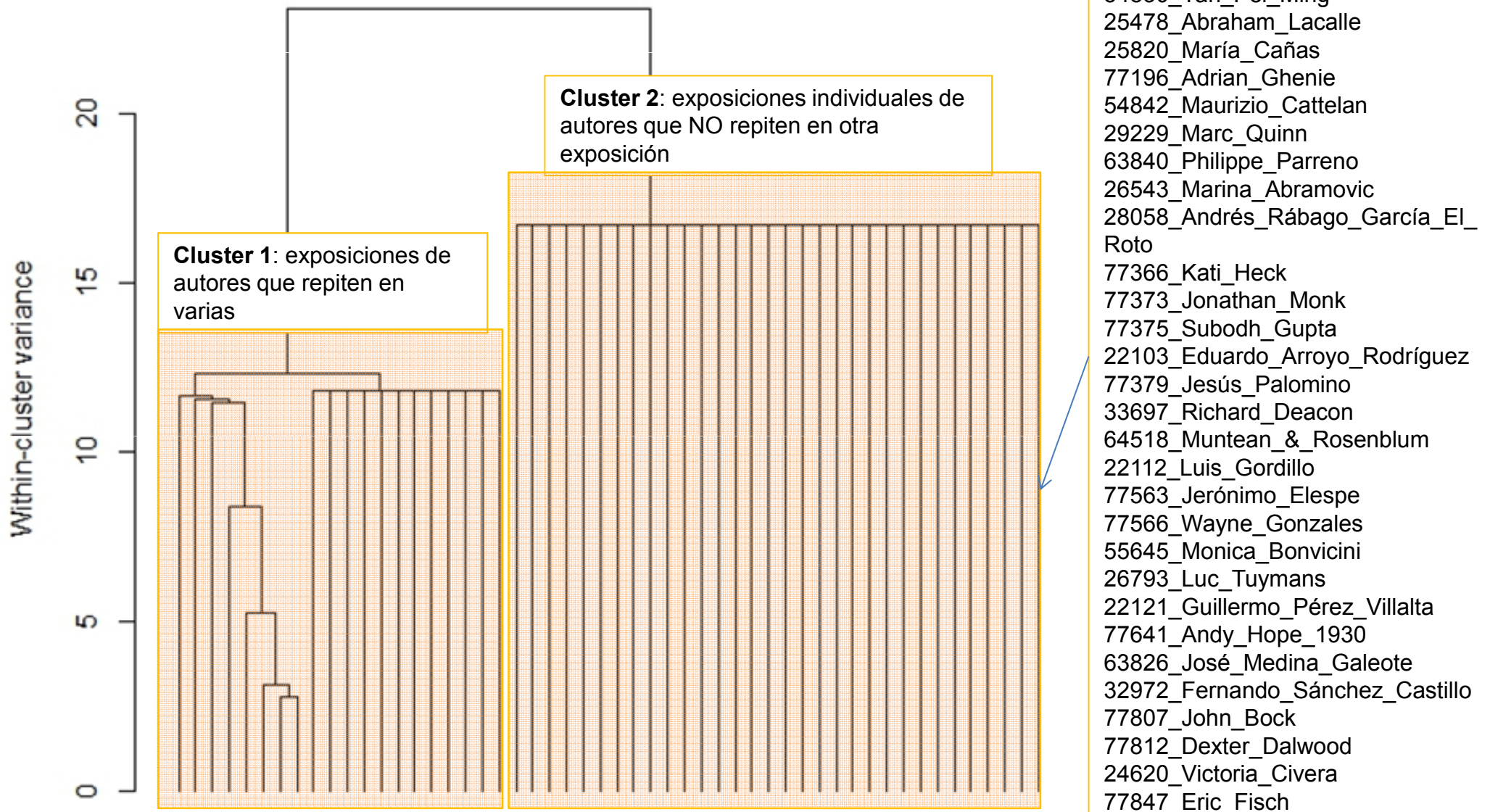


**SUBCORPUS CURADOR  
FERNANDO FRANCÉS**

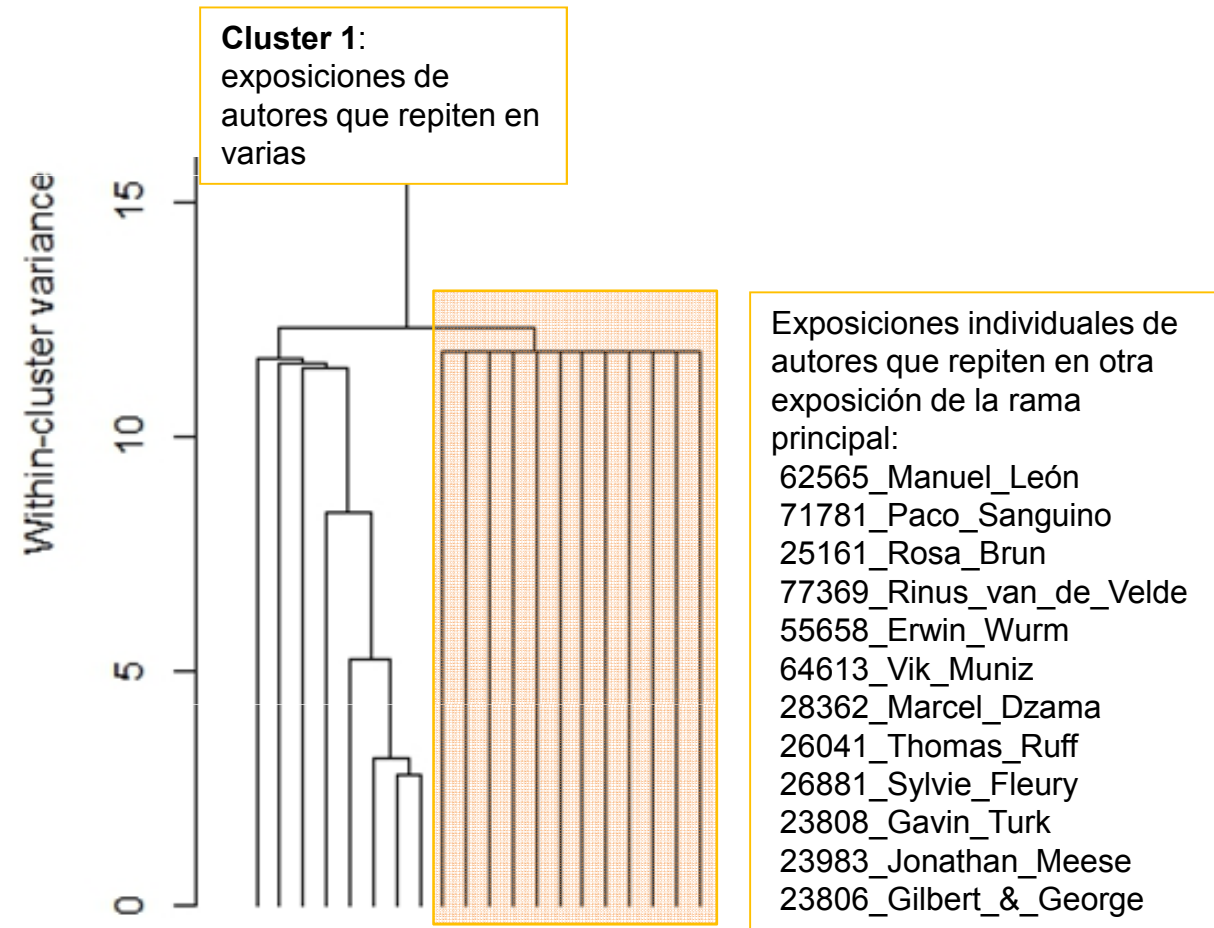


# Clasificación de las exposiciones del curador Fernando Francés

Full documents dendrogram



# Estudio del cluster 1, rama derecha



# Estudio del cluster 1, rama izquierda

**Cluster 1:**  
exposiciones de  
autores que repiten en  
varias

27286\_JR

27173\_D\*face

27092\_OBEY

27165\_Peter\_Saul 27166\_Erik\_Parker  
27172\_KAWS

27092\_OBEY 27172\_KAWS 27173\_D\*face 27286\_JR  
28298\_Banksy 77406\_Daniel\_&\_Geo\_Fuchs

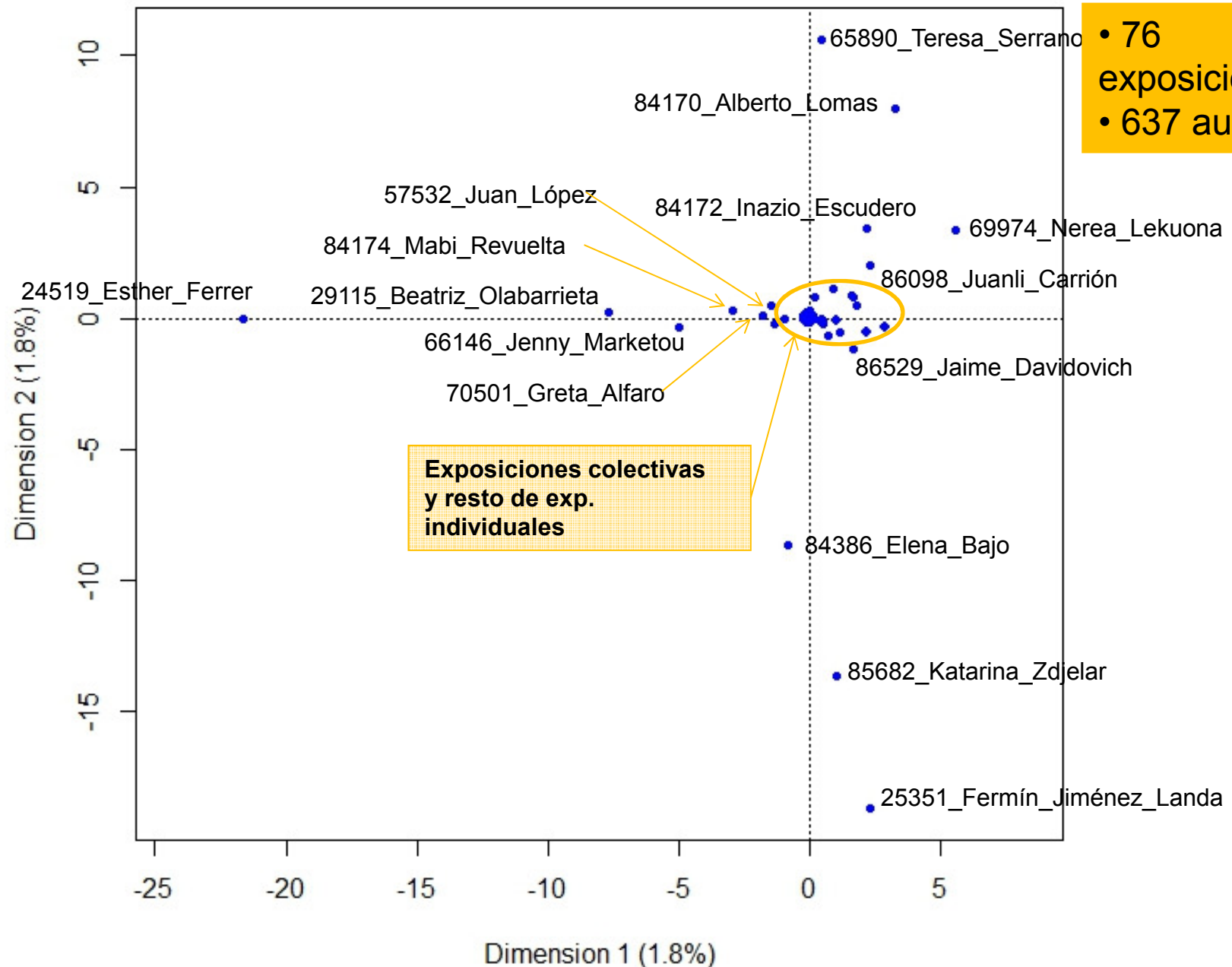
22105\_Miquel\_Barceló\_Artigues 22106\_Carmen\_Calvo  
22113\_Cristina\_Iglesias\_Fernández\_Berrido  
22133\_Susana\_Solano 23362\_Louise\_Bourgeois  
23917\_Jean\_Michel\_Basquiat  
24365\_José\_Manuel\_Broto\_Gimeno 24438\_Juan\_Muñoz  
25452\_Andy\_Warhol 25477\_Juan\_Uslé 25854\_Txomin\_Badiola  
25876\_Michel\_François 26879\_Ólafur\_Eliasson  
26881\_Sylvie\_Fleury 28378\_Jenny\_Holzer  
39545\_Susy\_Gómez 39575\_Chema\_Alvargonzález  
39598\_Eve\_Sussman 42288\_Bill\_Viola 51108\_Anish\_Kapoor  
63557\_Kimsooja 71832\_Matthew\_Ritchie 71833\_Rong\_Rong  
71834\_Julio\_Galán 71836\_Adam\_McEwen  
71838\_José\_María\_Cano 82008\_Günther\_Förg

23835\_Paco\_Aguilar 23847\_Sebastián\_Navas  
23906\_Carlos\_Aires 23941\_Javier\_Calleja  
25161\_Rosa\_Brun 25891\_Federico\_Guzmán  
32926\_Pilar\_Albaracín 33397\_Santiago\_Ydáñez  
40218\_Marina\_Vargas 46234\_Celia\_Macías  
46921\_Mari\_García 53081\_Chema\_Cobo  
54840\_Mónica\_Sánchez\_Robles  
57819\_Nono\_Bandera 62565\_Manuel\_León  
62573\_David\_Escalona 62589\_José\_Luis\_Puche  
64757\_MP&MP\_Rosado 71777\_Chema\_Lumbreras  
71781\_Paco\_Sanguino 71789\_Roy\_Anglada  
71793\_Miguel\_Gómez\_Losada 71796\_Jaro  
71801\_José\_Carlos\_Naranjo\_Bernal  
71805\_Juanma\_Reyes

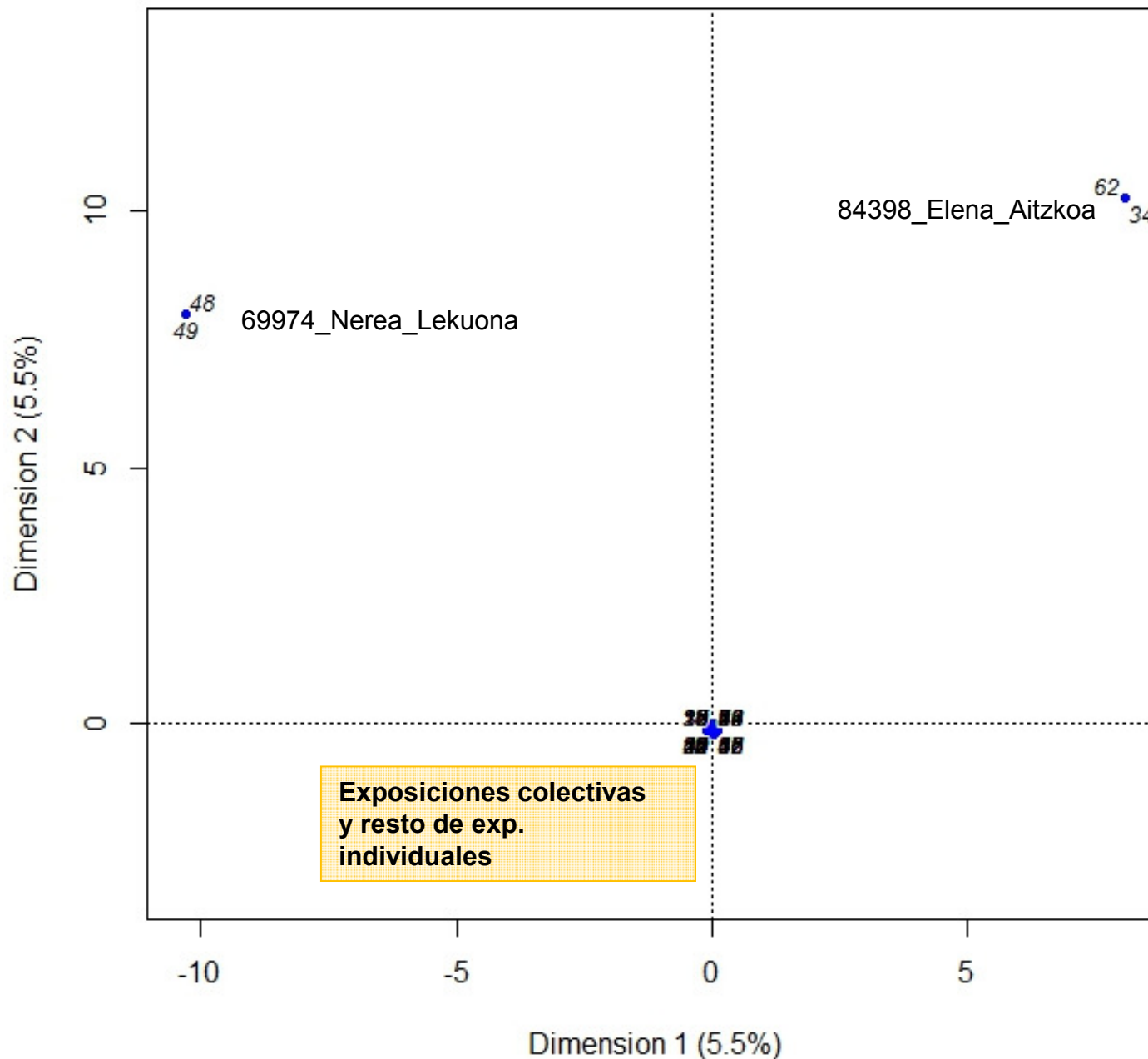
22105\_Miquel\_Barceló\_Artigues 23362\_Louise\_Bourgeois 23806\_Gilbert\_& George  
23808\_Gavin\_Turk 23937\_Tracey\_Emin 23981\_Paul\_McCarthy 23983\_Jonathan\_Meese  
24005\_Raymond\_Pettibon 24020\_Santiago\_Sierra 24455\_Gerhard\_Richter  
25906\_Thomas\_Hirschhorn 26029\_Julian\_Opie 26041\_Thomas\_Ruff  
26110\_Julião\_Sarmento 26548\_Rodney\_Graham 26581\_Lawrence\_Weiner  
26932\_Vanessa\_Beecroft 26973\_Liam\_Gillick 28362\_Marcel\_Dzama 29043\_Alex\_Katz  
33353\_Tony\_Cragg 39443\_William\_Kentridge 51108\_Anish\_Kapoor 53081\_Chema\_Cobo  
55658\_Erwin\_Wurm 64613\_Vik\_Muniz 68066\_Art\_and\_Language 69330\_Per\_Barclay  
69529\_Ilya\_Kabakov 77369\_Rinus\_van\_de\_Velde 77370\_Neo\_Rauch  
77371\_Jake\_&\_Dinos\_Chapman

**SUBCORPUS DE LA  
INSTITUCIÓN ARTIUM  
(MUSEO VASCO ARTE  
CONTEMPORÁNEO**

# Mapa de las exposiciones (autores mínima ocurrencia 1)

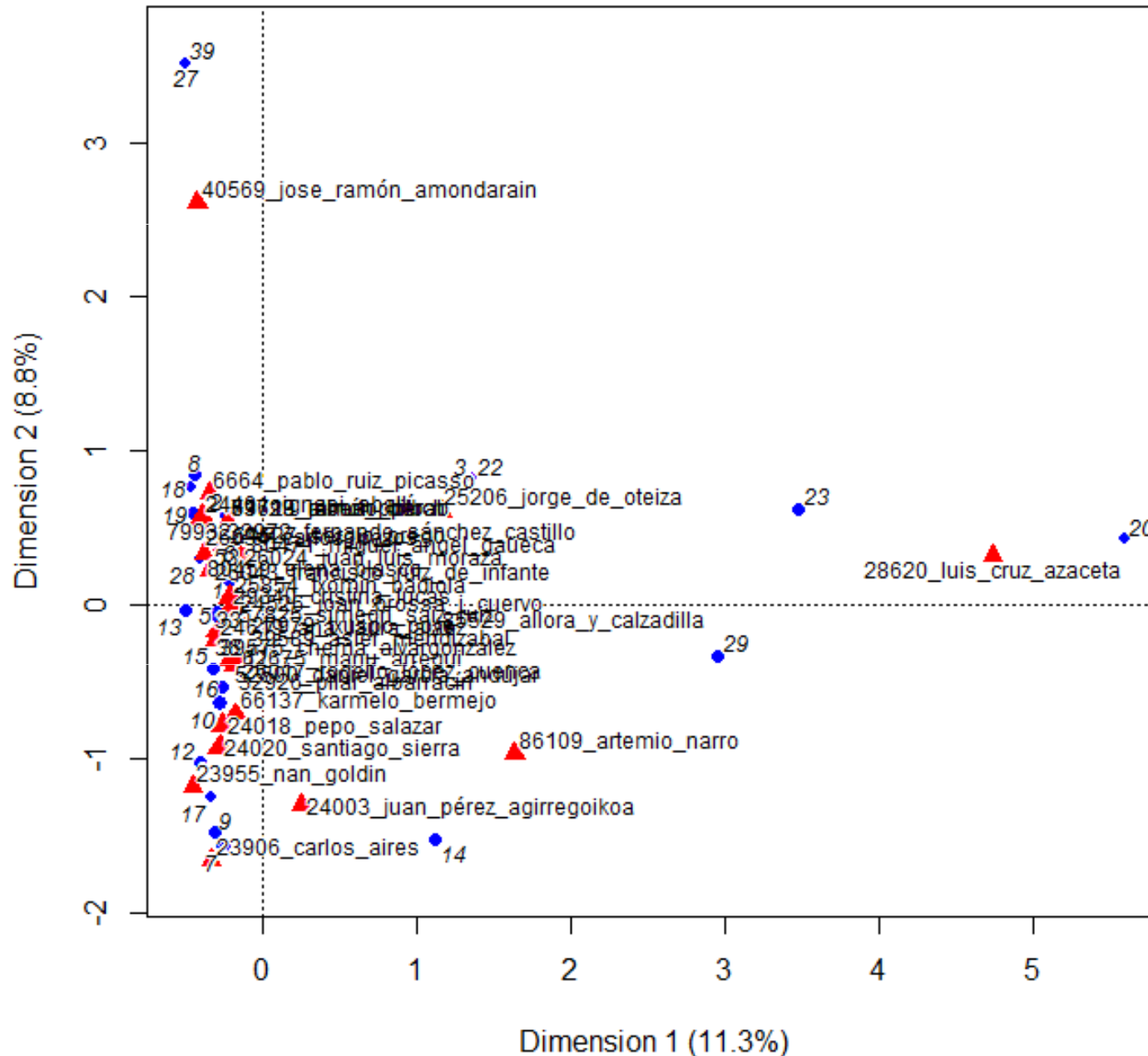


# Mapa de las exposiciones (autores mínima ocurrencia 2)



- 47 exposiciones
- 140 autores

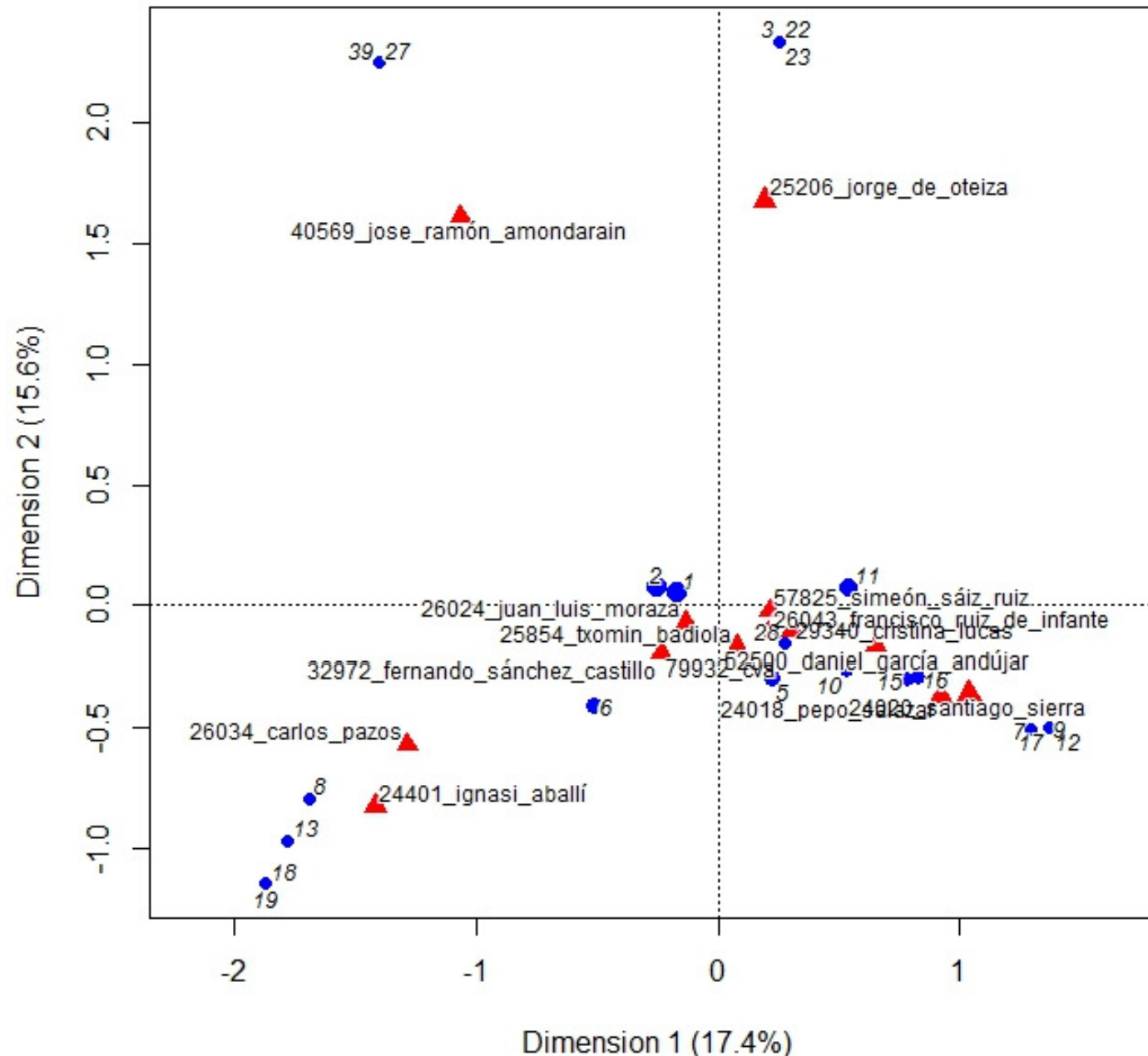
# Mapa de las exposiciones (autores mínima ocurrencia 3)



- 47 exposiciones
- 140 autores



# Mapa de las exposiciones (autores mínima ocurrencia 4)



- 22 exposiciones
- 14 autores



# **ANÁLISIS COMPLEMENTARIOS**

# Análisis estratégico (co-word analysys)

Parámetros red :

- Mín. Ocurr. 2
- Mín. Cooc. 2
- Mín. nº nodos 4
- Max. nº nodos 14

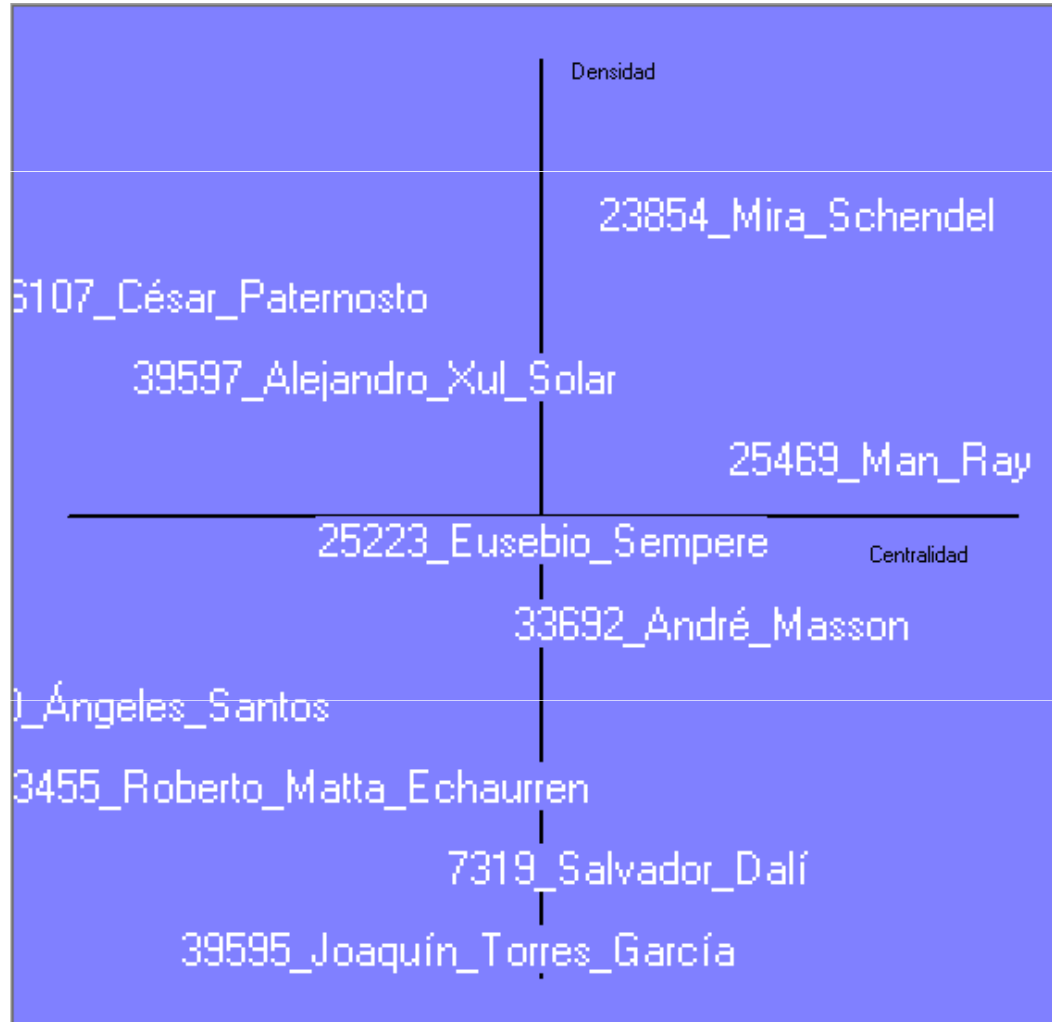
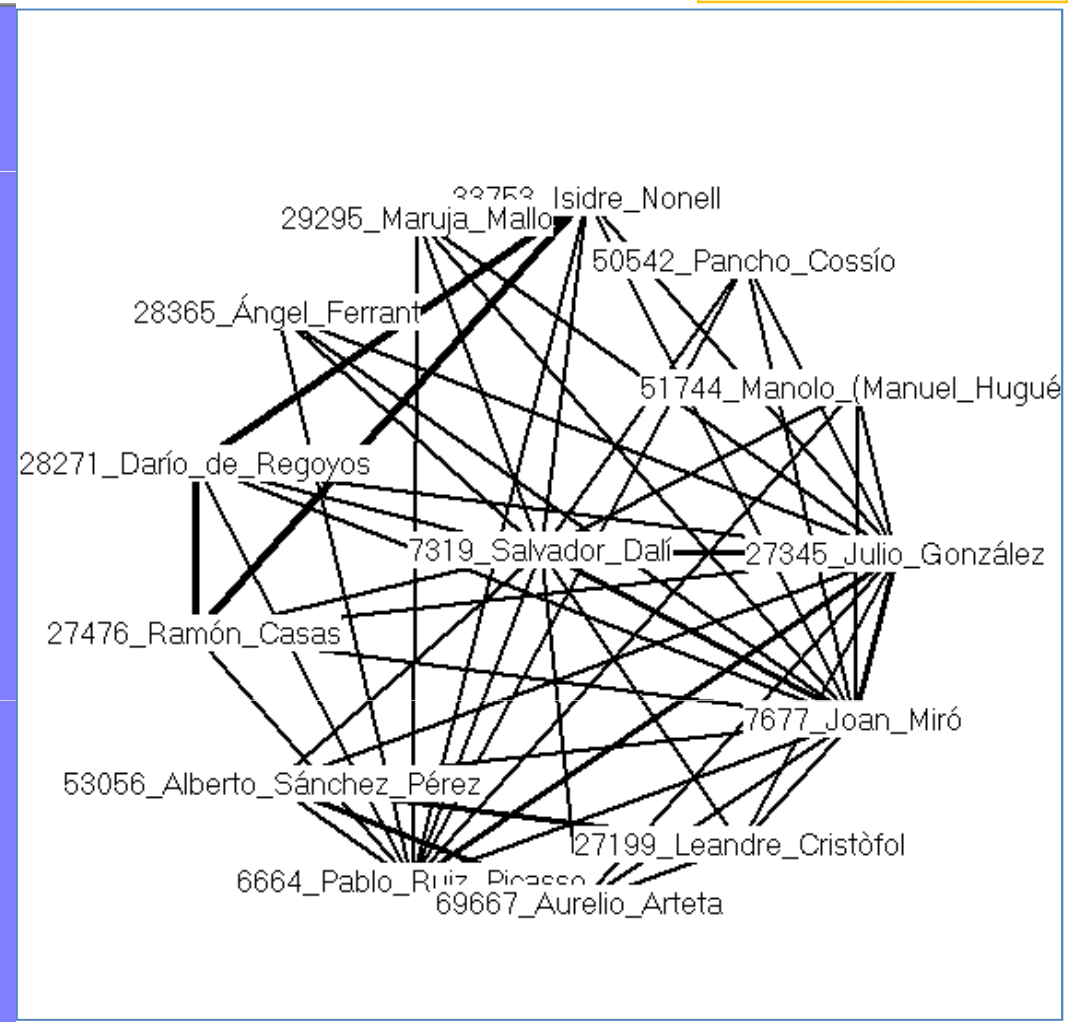
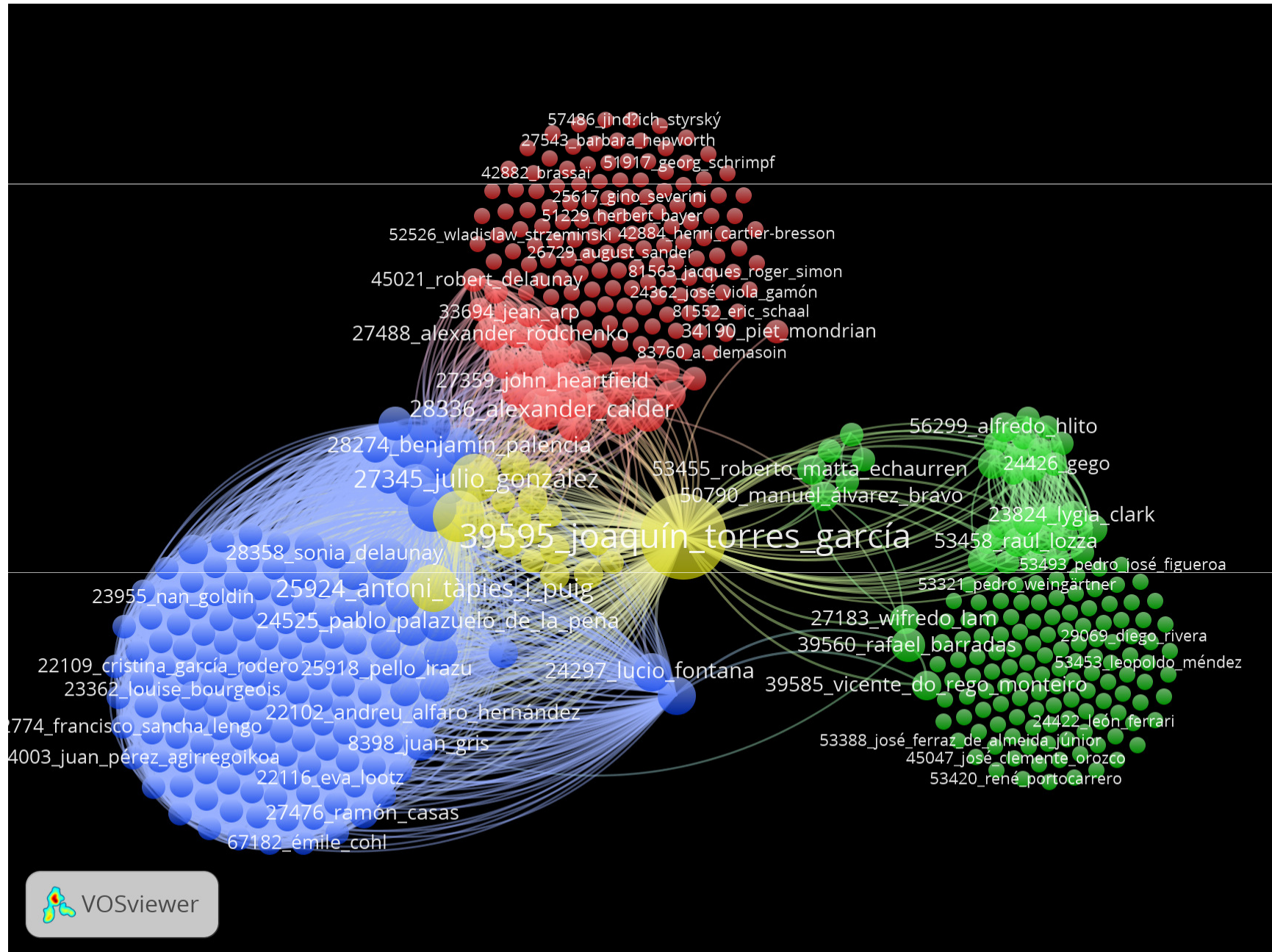


Diagrama estratégico



Grafo de la subred Salvador Dalí

# Visualización de redes (co-word analysis)



**TUTORIAL**

# Tutorial R.TeMiS (Pino, 2016)

2016

## Tutorial de R-Text Mining Solution



Prof. Dr. José Pino-Díaz  
Universidad de Málaga, Andalucía  
Tech, Escuela Técnica Superior de  
Ingeniería Industrial, Campus de  
Teatinos s/n, 29071 Málaga, España  
18/07/2016

### Contenido

Introducción .....	4
Comenzando a trabajar con R.TeMiS .....	5
Importar Corpus .....	8
Visualización del corpus activo y de los diccionarios de términos .....	13
Visualizar el corpus de documentos.....	13
Visualizar el diccionario de términos.....	14
Gestión y distribución del corpus.....	15
Análisis descriptivo del léxico.....	21
Resumen cuantitativo del vocabulario de términos. ....	22
Tabla de disimilaridad.....	23
Términos más frecuentes.....	24
Términos específicos por modalidades de la variable. ....	25
Análisis de términos concretos. ....	27
Términos que coocurren con otros concretos.....	29
Evolución temporal de términos concretos. ....	29
Análisis de correspondencias aplicado a un corpus de documentos .....	30
Concepto de similaridad, disimilaridad, distancia y proximidad entre documentos. ....	30
Análisis de correspondencias.....	30
Análisis de correspondencias con RTemis. ....	31
Interpretación del diagrama de correspondencias .....	34
Procedimiento de análisis de correspondencias con RTemis. ....	35
I) AFC de la matriz documentos-términos sin agregar ninguna variable ...	35
II) AFC de la tabla lexical completa agregando variables ( <i>full document-term matrix by variables</i> ).....	40
Clasificación ascendente jerárquica aplicada a un corpus de documentos .....	45
Concepto de similaridad y distancia entre documentos. ....	45
Concepto de clasificación jerárquica. ....	45
Clasificación ascendente jerárquica.....	46
Clasificación ascendente jerárquica con RTemis. ....	47

**MUCHAS GRACIAS**